1-1-1975

A generalized longitudinal factor model for the analysis of change.

James Algina
University of Massachusetts Amherst

A GENERALIZED LONGITUDINAL FACTOR MODEL FOR THE ANALYSIS OF CHANGE

A Dissertation Presented

By

James Algina

Submitted to the Graduate School of the
University of Massachusetts in partial
fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

November 1975

A GENERALIZED LONGITUDINAL FACTOR MODEL FOR THE ANALYSIS OF CHANGE

A Dissertation

By

James Algina

Approved as to style and content by:

Hari Haran Swaminathan

Dr. Hariharan Swaminathan, Chairman of Committee

Ronald K. Hambleton

Dr. Ronald K. Hambleton, Member

Dr. Alexander Pollatsek, Member

Alexander Pollatsek

Louis Fischer

Louis Fischer, Acting Dean

School of Education

November 1975

ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee, Drs. Hariharan Swaminathan, Ronald K. Hambleton and Alexander Pollatsek not only for their assistance throughout the dissertation, but also for the contributions each has made to my development throughout my graduate career.

I owe a special debt of gratitude to Dr. Swaminathan for working so closely with me on this dissertation and the other research we have collaborated on.

I also owe a special debt of gratitude to Dr. Hambleton for the incredible energy he expended to make the Laboratory of Psychometric and Evaluative Research an extremely supportive environment to work in.

Finally, I would like to express my appreciation to Dr. Daniel C. Jordan for serving as the Dean's representative at the oral defense and for his kind praise at the conclusion of the orals.

A GENERALIZED LONGITUDINAL FACTOR MODEL
FOR THE ANALYSIS OF CHANGE

James Algina

Abstract

Much of educational research and, for that matter, a significant proportion of all behavioral research is concerned with the study of behavioral change. Frequently, an investigator is interested in change in a number of variables and therefore collects multiple test data on each of several occasions. In this situation change can be interpreted either as change in the average level of performance on the tests or as change in the relationships among the tests. When the relationships among the tests are the foci of interest and when the tests are correlated, it is reasonable to assume that latent variables underlie performance on the tests. In this situation it is more meaningful to study the relationships among these underlying variables than to study the relationships among the tests. In addition, the study of the relationships among the tests and the latent variables will provide valuable information concerning the nature of change.

In many situations, the latent variables underlying performance on the test may be identified as the factors of the common factor model. The relationships among the latent variables may then be studied by testing hypotheses about the invariance of the factor scores. The relationships among the latent variables and

the tests may be studied by testing hypotheses about the invarinace of the factor patterns, factor structures and factor-test correlations.

In order to develop procedures for testing these hypotheses, a structural model, based on the assumption that the common factor model underlies the observations at each of two occasions, is postulated. This structural model is intended to permit simultaneous estimation of 1) the factor patterns for each occasion, 2) the first occasion factor score covariances, 3) the regression weights for the regression of the second occasion factors on the first occasion factors, 4) the unique factor score variances, and, 5) the between occasion unique factor score covariances. Since maximum likelihood estimates are asymptotically consistent, sufficient and efficient and because maximum likelihood estimation permits testing hypotheses about the estimators, maximum likelihood equations for estimating the parameters of the model are derived. In addition, the second derivatives of the likelihood function are provided to facilitate the utilization of the Newton-Raphson procedure for solving the likelihood equations.

The questions of the invariance of the common factor scores, the unique factor scores, the unique factor variances, the factor patterns, factor structures, and factor-test correlations as well as the question of the adequacy of the structural model should be investigated by testing structural hypotheses about the parameters of the structural model. Tests of hypotheses appropriate for examining these questions are constructed using the likelihood ratio criterion. In addition, expressions for the asymptotic variance covariance matrix of the maximum likelihood estimators are derived so that confidence intervals and regions for the parameters can be constructed.

Finally, two extensions of the model are briefly discussed. The first discussion concerns the problem of investigating the similarities and differences, in the structure of the two occasion covariance matrix, for independent groups. The second extension consists of the development of the structural equations for a k occasion longitudinal factor analysis model.

Table of Contents

CHAPTER I

INTRODUCTION

1.0 Statistical Procedures for the Study of Behavioral Change

Much of educational research, and, for that matter, a significant proportion of all behavioral research is concerned with the study of behavioral change. Naturally enough this concern has led to the development of several statistical procedures for the study of change. These procedures may be thought of as belonging to either of two classes. Methods belonging to the first class are intended to yield information about change in the level of behavior. Change in the level of behavior is usually interpreted as change in the mean of a measurement of that behavior. Procedures for assessing change in the mean level of behavior, that are applicable to both experimental and quasi-experimental research designs, and the problems associated with assessing change have been discussed by Algina and Swaminathan (1975), Bock (1963), Campbell and Stanley (1966), Glass, Wilson and Gottman (1972), Grizzle and Allen (1969), Lord (1958, 1963, 1967, 1969), O'Connor (1972), Pothoff and Roy (1964), Rao (1959, 1965, 1966, 1967), Timm (1974), Werts and Linn (1970), and Werts, Joreskog and Linn (1972), among many others.

The second class of statistical methods is comprised of techniques intended to yield information about change in the relationships among behaviors. Generally speaking, the purpose of such methods is to analyze the covariance or correlation matrix obtained from multi-measurement, multi-occasion data, in order to identify latent sources of variations, and study how these latent sources of variations change as a function of time and intervening treatment. Identifying latent sources of variations in this case is tantamount to isolating traits or characteristics of individuals that cannot be measured directly but which are assumed to affect suitably chosen variables that can be measured or observed. Questions dealing with change are represented by such questions as whether the underlying traits or characteristics tend to remain invariant, tend to become integrated, or tend to become more specific over time. Since factor analysis is a procedure designed to isolate latent sources of variation, one way of proceeding to answer such questions is to postulate that some sort of a factor analysis model is operating at each occasion, estimate the parameters of the model and study change in these parameters. Examples of such statistical methods are given by the works of Corballis (1972), Corballis and Traub (1970), Harris (1963), Joreskog (1969), Swaminathan (1972, 1973), and Tucker (1963). As these methods all employ the factor model, to some extent, they are referred to as longitudinal factor analysis procedures.

The importance of analyzing change in the level of behavior would probably be readily admitted by many educational researchers. However, since the analysis of change in the relationships among behaviors is less familiar than analysis of change in the level of behavior to most educational researchers, the importance of the former mode of analysis may not be recognized. In the next section, an attempt is made to argue that longitudinal factor analysis is indeed an important mode of data analysis. Also, since the longitudinal factor analysis procedures are likely to be less familiar to educational researchers than procedures for analyzing mean change, a review and critique of longitudinal factor analysis methods will be presented in section 1.2.

1.1 The Importance of Longitudinal Factor Analysis

It is tempting to try to support the importance of longitudinal factor analysis by arguing that a longitudinal factor analysis yields more important information than an analysis of change in the mean level of performance. However, the argument is fallacious as both types of procedures yield important information. The importance of longitudinal factor analysis arises, not because it yields better information than the other statistical methods, but rather because it yields information of a different nature. The information that factor analysis of longitudinal data can provide, is illustrated by the major questions the technique is intended to answer. These are the questions of change in the common and unique factor scores, change in the factor patterns and change in the unique variance of the measurements over the two occasions. The argument, that longitudinal factor analysis provides information of a different nature, can be strengthened by showing that the information is at least potentially useful information about the nature of change. To this end a hypothetical educational example is presented below.

Suppose that a set of measurements, depending on two verbal and two numerical factors, is made at the beginning and the end of an instructional sequence. If the factors are thought of as generating the observed variables, then both the within and between occasion correlations for the factors should provide information that is useful in developing a description of the change that takes place over the course of instruction. For a particular factor, the between occasion correlation indicates the degree to which the examinees' deviation factor scores remain the same over the instructional period. Therefore, these correlations should be important descriptors of change. The within occasion correlations indicate the nature of the interrelationships among the factors and change in the within occasion factor correlation matrix indicates how the factor interrelationships change over the course of instruction. Thus a comparison of the within occasion correlation matrices should provide some information about change.

Let us assume that the following pattern of data is observed. A longitudinal factor analysis of the pre and post-instructional measurements indicates that the within occasion correlation between the numerical factors increases over the two occasions and that the between occasion correlation for each numerical factor is about .50. On the other hand, the analysis of the verbal factors indicates that over the two occasions the within occasion correlation remains relatively stable. Furthermore, for each verbal factor there is a larger correlation between occasions than for the numerical factors.

These results provide a description of the pattern of change which is a rich basis for evaluating instruction. The increase in the within occasion correlation for the numerical factors suggests that the abilities represented by these factors become more highly related over the course of instruction. This may mean that, to a larger extent at the end of instruction than at the beginning, the numerical factors reflect a common core of knowledge and skills. Alternatively, it may be that the factors continue to represent different abilities, but the abilities have become more highly correlated by the end of instruction. If we assume that with most instructional methods the between occasion correlations for the numerical factors are substantially larger than .50, then the moderate between occasion correlations could be an indication that the instruction has tended to equalize the chance of success for those pupils who initially did well and for those who initially did poorly. In fact, if the between occasion correlations for the unique factors of the numerical tests are close to zero, and if the cross occasion relationships between the latent variables underlying the test performance are truly linear then this interpretation appears to be the only plausable interpretation of the data. It should be noted that the assumption that the between occasion correlations are substantially larger for other instructional methods could be investigated by conducting longitudinal factor analyses of data gathered on students exposed to other instructional methods.

On the other hand the results concerning the substantial between occasion correlations for each of the verbal factors suggest that the instruction has merely maintained the initial status of the students. The lack of change in the within occasion correlation for the verbal factors implies that the relationship between the abilities underlying the observed tests has not been changed by the instruction. However, in order to determine that these effects are related to the particular instruction, it is necessary to conduct longitudinal factor analyses on data gathered on students exposed to other instructional methods.

The description of the instructional effects given above is characterized by greater detail than would have been possible if only average levels of performance had been studied. This increased detail may make it possible to review instruction and develop hypotheses about why the results occurred. For instance, suppose a record of the instruction had been maintained and included information about the organization of the instruction and the extent to which students were encouraged to exercise the concepts involved in the instruction. A review of this record reveals that the numerical portion of the instruction was organized with many horizontal and vertical links between subtopics and was characterized by a great deal of recycling through material as a method of review. In addition, there was extensive opportunity to apply the concepts that were taught. The verbal portion of the instruction was organized into many discrete units with little or no linkage between the units. This review combined with the description of the outcome of instruction probably suggests that the verbal portion of instruction should be reorganized along the lines of the numerical portion of the instruction.

From this hypothetical example it may be seen that the methodology has the potential for providing useful information about change. The information is valuable in the sense that the pattern of change is described by the information. Further, this description of the pattern of change may suggest instructional improvements, and if it does some practical benefit can be attributed to the information.

Another argument for the importance of longitudinal factor analysis is based on the contention of Harris (1963) that an important area of psychometric investigation is the study of measurements as the conditions of measurement are varied. Such a study may be expected to provide insight into the measurements, that cannot be achieved by studying measurements obtained under a single measurement condition. Since, historically, one of the principal purposes of factor analysis has been to analyze measurements that are not well understood, it seems particularly appropriate to develop procedures for factor analyzing measurement data obtained under different conditions. Longitudinal factor analysis, of course, is just such a procedure and, therefore, the procedure should be useful in measurement research.

As an example, suppose that in a particular developmental study the factor scores remain relatively stable, and that a particular test has a large loading on one factor and a small loading on a second factor at the beginning of the study. Further, suppose that at the end of the study the situation is reversed. This result may indicate that the test requires different abilities at different stages of development. Although this interpretation is quite compelling, if the factors are correlated, it is important to examine tha factor structure matrices before making a decision about the interpretation. It should be noted that with correlated factors, the factor loadings are analogous to multiple regression weights for the regression of the observed scores on the factor scores, while the elements of the factor structure matrices are the covariances between the observed scores and the factors. Therefore, it is entirely possible for the factor loadings to change, while the elements of the factor structure matrices remain fairly stable. However, if the correlations between the test and the factors remain stable then the statement, that the abilities underlying the test have changed, is not justified. Stable correlations between the test and the factors imply that the relations between the tests and the factors have not changed and, hence, that the abilities underlying test performance remain the same. Suppose that after inspecting the factor structure matrices, the changing ability interpretation is still justified. If the abilities underlying performance on the two occasions can be identified, the results will suggest the age range in which the test is useful as a measure of each ability. These results certainly represent more information than would have been developed by a single occasion factor analysis.

1.2 A Review of Longitudinal Factor Analysis Methods

The earliest work in the field of longitudinal factor analysis proceeded in either of two ways. One procedure was to subject the entire correlation matrix to a common factor analysis. Letting $\underline{y}' = [\underline{y}_1' \; \underline{y}_2']$ denote the (1x2p) random vector of observations with $\underline{y}_1$, $\underline{y}_2$ denoting the first and second occasion random vectors of observations, respectively, the factor model on which this method is based may be written as

$$(1.1) \qquad \underline{y} = \Lambda\underline{x} + \underline{e} \quad ,$$

$$(1.2) \qquad = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} \underline{x} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \end{bmatrix} ,$$

where Λ_1 and Λ_2 are the (pxr) first and second occasion factor pattern matrices, respectively, $\underline{x}$ is the (rx1) random vector of factor scores, and $\underline{e}_1$ and $\underline{e}_2$ are the (px1) random vectors of first and second occasion unique scores. From (1.2) it follows that

$$(1.3) \qquad \underline{y}_1 = \Lambda_1\underline{x} + \underline{e}_1 ,$$

and

$$(1.4) \qquad \underline{y}_2 = \Lambda_2\underline{x} + \underline{e}_2 .$$

The factor score vector $\underline{x}$ is thus not defined separately for the two occasions. This has prompted Corballis and Traub (1970) to criticize the method on the grounds that a tacit assumption of factor score invariance is made. Corballis and Traub (1970) point out that the assumption of invariant factor scores implies that there are factors that represent immutable qualities of people. They further suggest that such immutable qualities probably do not exist. However, even if some such immutable qualities do exist, it is clear that not all factors represent such qualities. Thus, in any particular situation the invariance of factor scores is more appropriately treated as a hypothesis to be tested than as an assumption in the model.

The linear model (1.1) in conjunction with the assumptions of common factor analysis lead to the following equation for the structure of the correlation matrix

$$(1.5)\qquad \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} [\Lambda_1' \; \Lambda_2'] + \begin{bmatrix} \psi_{11} & 0 \\ 0 & \psi_{22} \end{bmatrix} ,$$

where the matrices ψ_{11} and ψ_{22} are diagonal. Examination of Equation (1.5) reveals that the unique factor scores are assumed to be uncorrelated over occasions. Since unique factors are hypothesized sources of individual differences, it seems unreasonable to assume that the unique factors for the same test do not covary over time. If they were uncorrelated over time it would probably be more appropriate to consider the unique factors as measurement error than as sources of individual differences.

Another point of view about the nature of the common factors is, in the language of Joreskog (1969), that they represent some combination of general factors, test specific factors and occasion specific factors. Test specific factors are defined as factors that contribute to the covariance between the scores on a given variable at different occasions; occasion specific factors are factors that contribute to the covariance between variables within a given occasion; and finally, general factors are those which contribute to covariances between the variables both within and between occasions. However, since attention has not been paid to defining these entities, it is difficult to sort out which factors fall into which category and, therefore, the results may not be interpretable.

The second method that earlier workers adopted was to perform a separate common factor analysis of the matrices R_{11} and R_{22} and rotate the factors to congruence for comparisons. While this approach, unlike the previous approach which permitted only varying factor patterns, would clearly permit varying factor scores and patterns, it fails to take into account information contained in the across occasion covariance matrix. The method also lacks an explicit statement of the assumed relationship between factors for the two occasions.

Harris (1963) presented an application of the canonical factor model to the problem of longitudinal factor analysis. In this application, the canonical factor model is based on (1.3) and (1.4) and consequently Equation (1.7). Thus Harris' approach may be subject to the same criticisms as those that were made about the method of common factor analyzing the entire correlation matrix; i.e., Harris' approach may be viewed as restrictive in the sense that factor scores are assumed to be invariant.

Tucker (1963) presented an application of the three mode factor model to the analysis of multi-measurement multi-occasion data. The basic equation of the model is

$$y_{ijk} = \sum_{mpq}\sum\sum a_{im} b_{jr} c_{kq} g_{mrq} \quad .$$

The indices i, j, and k refer to individuals, variables and occasions respectively. The model may be interpreted in the following fashion. The quantity g_{mrq} is an entry in a three dimensional array, G, with row dimension m, column dimension r, and depth dimension q. Each of the q matrices of dimension (mxr) in G contains the scores of m idealized individuals on r idealized traits for a particular idealized occasion. For each of the k occasions a (mxp) matrix N_k with general element n_{mr} is derived as a weighted sum of the q idealized occasion matrices. That is

$$n_{mr} = \sum_q c_{kq} g_{mrq} \quad .$$

The elements of the matrix N_k represent the scores of the m idealized individuals on the r idealized traits for a particular occasion. For each occasion the scores of the examinees on the r idealized traits are derived as weighted sums of the scores of the m idealized individuals on the r idealized traits. Let the number of examinees be N. Denote the (Nxr) matrix of the scores of the N individuals on the r idealized traits, with general element s_{ir}, as S_k. Then

$$s_{ir} = \sum_m a_{im} n_{mr} \quad .$$

In terms of matrices

$$S_k = AN_k \quad .$$

The matrix S_k' is analogous to a matrix of factor scores for occasion k.

From this, it may be seen that the factor score analogues, S_k', are permitted to vary over occasions. Finally the scores of N individuals on the p observed on the kth occasion are derived by the following equation:

$$y_{ijk} = \sum_r b_{jr} s_{ir} \quad ,$$

$$Y_k = BS_k' \quad .$$

The matrix B is analogous to the factor pattern matrix and from this equation it is clear that the model holds the factor loading analogues constant over the k occasions. Thus any description of change must be in terms of S_k', the factor score analogues. Hence, the three mode factor model may be viewed as holding the factor patterns invariant while permitting the factor scores to vary.

Thus, the early attempts at factor analyzing longitudinal data suffered from several drawbacks. In some methods either the factor scores or the factor patterns were held invariant, while in other methods both factor patterns and factor scores were permitted to vary, but the information contained in the cross occasion correlation or dispersion matrix was ignored. In addition, some of these methods assumed that the unique scores for the same variables were uncorrelated over occasions. Corballis and Traub (1970), in order to remedy this situation, presented a longitudinal factor model that would permit varying factor scores, varying factor patterns and would also permit correlated unique scores. In addition, the model takes into account the information contained in the cross occasion correlation or dispersion matrix.

Corballis and Traub (1970) assumed that

(1.6) $$\underline{y}_1 = \Lambda_1 \underline{x}_1 + \underline{e}_1 \quad ,$$

and

(1.7) $$\underline{y}_2 = \Lambda_2 \underline{x}_2 + \underline{e}_2 \quad .$$

In addition they assumed that

(1.8) $$E(\underline{x}_i \underline{x}_i') = I \quad , \quad i=1,2 \quad ,$$

$$E(\underline{x}_i \underline{x}_j') = 0 \quad , \quad i \neq j \quad ,$$

and

$$E(\underline{e}_i \underline{e}_j') = \psi_{ij} \quad , \quad i,j=1,2 \quad ,$$

with ψ_{ij} diagonal. They further assumed that the regression of $\underline{x}_2$ on $\underline{x}_1$ is given by the model

(1.9) $$\underline{x}_2 = D\underline{x}_1 + \underline{d} \quad ,$$

where D is a (rxr) diagonal matrix of regression weights and $\underline{d}$ is a (rx1) vector of residuals such that

$$E(\underline{d}\underline{d}') = \theta \quad , \quad \text{diagonal}$$

and they assumed that

$$E(\underline{x}_1\underline{d}') = 0 \quad .$$

Thus,

$$(1.10) \qquad \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} I & D \\ D & I \end{bmatrix} \begin{bmatrix} \Lambda_1' & 0 \\ 0 & \Lambda_2' \end{bmatrix} + \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix} \quad ,$$

where $\Sigma_{ij} = E(\underline{y}_i\underline{y}_j')$, and Σ_{11}, Σ_{22}, $\Sigma_{12}(=\Sigma_{21}')$ are the first occasion, second occasion, and cross occasion dispersion matrices respectively.

The work of Corballis and Traub, although a substantial improvement over the earlier attempts at longitudinal factor analysis, is not free from drawbacks. One weakness is the restriction to two occasions. Corballis (1973) and Swaminathan (1972, 1973) extended the model to accommodate more than two occasions. They assumed that

$$(1.11) \qquad \underline{y}_i = \Lambda_i\underline{x}_i + \underline{e}_i \quad ,$$

with

$$(1.12) \qquad E(\underline{x}_i) = \underline{0} \quad ,$$

$$(1.13) \qquad E(\underline{x}_i \underline{x}_i') = I \quad ,$$

$$(1.14) \qquad E(\underline{e}_i) = \underline{0} \quad ,$$

$$(1.15) \qquad E(\underline{e}_i \underline{e}_j') = \psi_{ij} \quad , \quad i,j=1,\ldots,k,$$

and that the regression of $\underline{x}_i$ on $\underline{x}_{i-1}$ is given by

$$(1.16) \qquad \underline{x}_i = D_{i-1}\underline{x}_{i-1} + \underline{d}_i \quad , \quad i=1,\ldots,k,$$

where

$$(1.17) \qquad E(\underline{d}_i) = \underline{0} \quad ,$$

$$(1.18) \qquad E(\underline{d}_i \underline{d}_i') = \theta_i, \quad \text{diagonal} \quad ,$$

and

$$(1.19) \qquad E(\underline{d}_i \underline{d}_j') = 0 \quad ,$$

$$(1.20) \qquad E(\underline{d}_i \underline{x}_j') = 0 \quad , \quad i,j=1,2,\ldots,k, \quad i>j \quad .$$

Assumptions (1.16), (1.19) and (1.20) imply that

$$(1.21) \qquad E(\underline{x}_i \underline{x}_j') = D_i D_{i+1} \cdots D_{j-1} \quad , \quad i<j \quad .$$

A problem with the works of Corballis and Traub (1970) and Corballis (1973) is that their procedures for the estimation of parameters were unsatisfactory. As Swaminathan (1972, 1973) pointed out, the two-stage Least-Squares procedure adopted by Corballis and Traub (1970) does not necessarily yield the true Least-Squares estimators, while Corballis (1973) acknowledges that his procedure for the estimation of parameters had an "improvised quality". Moreover, Corballis (1973) estimated the inter-occasion factor co-variance matrices under the tacit assumption that they are mathematically independent of one another, an assumption which by virtue of (1.21) is not true. As a result, constrained parameters are estimated as if they were free parameters. This may lead to inconsistent estimates of the parameters. The most serious weakness of the works of Corballis and Traub (1970) and Corballis (1973) is that they did not provide statistical tests of significance for the various hypotheses of interest. As only sample estimates of the population parameters are obtainable, the question of parametric change across occasions is an issue in statistical inference.

Swaminathan (1972, 1973) provided methods for obtaining maximum likelihood estimates of the parameters, incorporating in the procedure a proper treatment of constrained and free parameters. In addition, he derived likelihood ratio tests for testing hypotheses of interest and introduced the notion of factorial invariance over occasions.

Swaminathan's model (1972, 1973), like the models of Corballis (1973) and Corballis and Traub (1970), assumes that factor scores are orthogonal. The assumption of orthogonal factors implies that the between occasions factor score covariance matrix should be diagonal. As a result, the model may be derived by application of canonical correlation to the factor scores. Since canonical correlation produces pairs of linear combinations, for which the within pair correlations are maximized and all between pair correlations are zero, the model developed by Corballis and Traub (1970) and, hence, that of Corballis (1973) and Swaminathan (1972, 1973) may be viewed as maximizing the temporal stability of the factor scores (Nesselroade, 1972). As Nesselroade (1972) pointed out, this may not be desirable in all contexts. In addition, as a result of factor orthogonality, the factor pattern matrices are uniquely determined and, hence, cannot be rotated to yield a more meaningful solution. Thus, the solution may not be easily interpretable. For these reasons the assumption of orthogonality seems quite restrictive. Furthermore, there is a general argument against the assumed orthogonality of factor scores that apparently was originally raised by Thurstone (1947). The assumption of orthogonality implies that the factors are uncorrelated. If a maximum likelihood solution is sought and the multivariate normality assumption of the procedure is taken seriously, the orthogonality assumption implies that the factors are statistically independent. If the factor model is viewed as a serious psychological model, that is, if the factors are viewed as the important sources of individual differences, then orthogonal factors imply that these important traits are in general either uncorrelated or statistically independent in in the population. This would seem to be a hypothesis rather than an assumption to be built into a model. It appears that the only way to overcome the limitations due to orthogonality is to formulate a model that permits correlated factors.

Joreskog (1969) proceeding along different lines presented an application of his general model for the analysis of covariance structures (Joreskog, 1970) to the problem of longitudinal factor analysis. In this application Joreskog (1969) postulated three kinds of factors to account for the covariance or correlation matrix of multi-measurement multi-occasion data. As was mentioned earlier, these are test specific factors which contribute to the covariance between the results of a given test on different occasions; occasion specific factors which contribute to covariances between tests within a given occasion; and general factors which contribute to covariances between the tests, both within and between occasions.

These definitions, which were given by Joreskog (1969) need to be amplified. General factors do contribute to covariances between tests, both within and between occasions, but, in addition, are defined to be invariant over occasions. Only tests for occasion t may load on an occasion specific factor for occasion t and in this sense, these factors contribute only to within occasion covariance. However, in at least one of Joreskog's models occasion specific factors appear to covary over occasions and in this sense contribute to inter-occasion covariance. In a similar vein only one test is permitted to load on each test specific factor, however test specific factors are permitted to covary with one another in several of Joreskog's models.

The model given by Joreskog (1969) for the measurements, $\underline{y}_t$, at time t may be written as

$$(1.22) \qquad \underline{y}_t = \underline{\mu}_t + A_t\underline{f} + B_t\underline{g}_t + C_t\underline{h} + \underline{e}_t \quad ,$$

where $\underline{y}_t$ is the (nxl) vector of observed scores, $\underline{\mu}_t$ is the mean vector of $\underline{y}_t$, $\underline{f}$(pxl) is a vector of general factor scores, $\underline{g}_t(q_t\text{xl})$ is a vector of factors specific to occasion t, $\underline{h}$(nxl) is a vector of test specific factors and $\underline{e}_t$(nxl) is a vector of unique factor scores. The matrices A_t(nxp), $B_t(\text{nx}q_t)$, and the diagonal matric, C_t(nxn), are factor loading matrices. By the exclusion of one or more of the vectors $\underline{f}$, $\underline{g}_t$, $\underline{h}$ and/or by making different assumptions about the covariances within and between the elements of the three vectors, a number of different longitudinal factor analysis models can be obtained. The important point to note is that the types of factors that are assumed to be operating and the inter-relationships among these factors are defined explicitly and, hence, can be studied.

One criticism of Joreskog's work is that although different factor models could be postulated by the specification of $\underline{f}$, $\underline{g}_t$, or $\underline{h}$, all of his models appear to be special cases of a general model for longitudinal factor analysis. The general model may facilitate testing hypotheses in a manner that can lead to the most appropriate, statistically speaking, combination of the three types of factors. Joreskog's approach by postulating a series of separate models may obscure this combination. On the other hand, the notion of different kinds of factors is theoretically important and some of the separate models suggested by Joreskog (1969) may not have become apparent through an inspection of the more general model. One such example is his model with general factors only. This model may be used to test hypotheses about whether tests given on different occasions are parallel, tau-equivalent, congeneric or otherwise.

A second criticism concerns one particular model postulated by Joreskog (1969). This is a model with occasion specific factors only. The regression of $\underline{g}_i$ on $\underline{g}_{i-1}$ is assumed to be given by

$$\underline{g}_i = D_{i-1}\underline{g}_{i-1} + \underline{d}_i \quad ,$$

with D_{i-1} diagonal. This model is the same as the model given by (1.6), however, Joreskog (1969) does not restrict

$$E(\underline{d}_i\underline{d}_i') = \theta_i$$

to be a diagonal matrix. Since the matrix θ_i may be non-diagonal, the factors are permitted to covary over occasions. The problem that arises is that since the factors are permitted to covary between occasions, unless the matrix of multiple regression weights for the regression of $\underline{g}_i$ on $\underline{g}_{i-1}$ is in fact diagonal, the model will not fit the observed covariance matrix very well. In fact, in analyzing some data by this model, Joreskog (1969) did find a poor fit to the data and "... found it difficult to improve the fit and still retain a meaningful interpretation".

One last problem in Jöreskog's work is that it is necessary to make the assumption that the unique score factors do not covary over occasions. However, to be fair, in the models in which the test specific factors for different tests are not permitted to covary, the inter-occasion factor covariances for these test specific factors may be the same as the unique score covariances.

1.3 Purposes of the Investigation

On the basis of the arguments given in section 1.1, the potential importance of longitudinal factor analysis was established. From the discussion in section 1.2, it may be argued that additional work on longitudinal factor analysis is merited. In addition, it is possible to abstract from the discussion in section 1.2 some important provisions that should be incorporated in a longitudinal factor analysis model. These provisions are:

1. varying common factor scores,
2. varying factor patterns,
3. specific factor scores that are correlated over occasions,
4. oblique factors,

and

5. statistical tests of significance for the relevant hypothesis.

The problem addressed in this dissertation is the development of a factor model, valid for two occasions, that incorporates these five provisions.

1.4 Outline of the Dissertation

The remainder of the dissertation is presented in four chapters. Chapter two presents the structural model and examines the questions of existence and identification of the model. In chapter three, the likelihood equations for estimating the parameters are given, and methods of solving the likelihood equations are briefly reviewed. Next, the second derivatives of the likelihood function, which are required for one of the methods of estimating the parameters, and also for deriving the asymptotic standard error of the estimates, are given. The effect of changes in the observed score metric on the parameters and parameter estimates are then discussed. In chapter four, hypothesis testing methods are presented. In the last chapter, the problem of studying similarities and differences in the two occasion longitudinal factor model, for different groups, is examined. Then the structural equations for a k occasion longitudinal factor model are developed. Finally, the limitations of the study are discussed.

CHAPTER II

A LONGITUDINAL COMMON FACTOR MODEL

2.0 Introduction

This chapter presents a longitudinal common factor model for analyzing multi-response data collected for the same individuals on two occasions. In section 2.1, the linear equations and the assumptions that underlie the structural model are given. From the equations and assumptions, the structural equations are derived. In section 2.2, the issues of existence and identification of the solution to the structural equations are discussed. In section 2.3, the relation between the conditions required for identification and the conditions required for solution of the likelihood equations is explored. The important distinction between arbitrary and theoretically based restrictions on the structure is made. Finally, in section 2.4, the possible effect of the conditions for identification and/or estimation on hypothesis testing is examined briefly, by presenting two examples of such effects.

2.1 The Model

Within each of the two occasions, it is assumed that the common factor model holds, i.e.,

$$(2.1) \qquad \underset{\sim}{y}_i = \Lambda_i \underset{\sim}{x}_i + \underset{\sim}{e}_i \quad , \quad i=1,2 \quad ,$$

where

$\underline{y}_i$ is the (px1) random vector of observed scores for the ith occasion,

Λ_i is the (pxr) matrix of factor loadings for the ith occasion,

$\underline{x}_i$ is the (rx1) random vector of factor scores for the ith occasion,

and

$\underline{e}_i$ is the (px1) random of unique scores for the ith occasion.

In addition, without loss of generality, it is assumed that

(2.2a) $E(\underline{y}_i) = \underline{0}$,

(2.2b) $E(\underline{e}_i) = \underline{0}$,

and

(2.2c) $E(\underline{x}_i) = \underline{0}$.

Furthermore, it is assumed that

(2.3a) $E(\underline{x}_i\underline{e}_j') = 0$, $i,j=1,2$,

and

(2.3b) $$E(\underline{e}_i\underline{e}_j') = \psi_{ij} = \psi_{ji} \quad , \quad i,j=1,2 \quad ,$$

where ψ_{ij} is diagonal. Assumption (2.3a) states that factor scores are orthogonal to unique scores both within and between occasions. In addition, assumption (2.3b) indicates that unique scores on different tests are orthogonal within and between occasions while unique scores on the same test may be correlated over occasions.

In order to formulate a model for changing factor scores, it is assumed that the regression of factor scores on the second occasion, $\underline{x}_2$, on $\underline{x}_1$, the factor scores on the first occasion, is given by

(2.4) $$\underline{x}_2 = D\underline{x}_1 + \underline{d} \quad .$$

The matrix D is an (rxr) matrix of regression weights, while $\underline{d}$ is the (rx1) random vector of residuals. The residuals are subject to the following assumptions:

(2.5a) $$E(\underline{d}) = \underline{0} \quad ,$$

(2.5b) $$E(\underline{d}\underline{x}_1') = 0 \quad ,$$

(2.5c) $$E(\underline{d}\underline{d}') = \theta \quad ,$$

where θ(rxr) is not necessarily diagonal.

Equations (2.1) for the two occasions may be combined into a single equation as follows:

$$(2.6)\qquad \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \end{bmatrix} ,$$

or, more compactly,

$$(2.7)\qquad \underline{y} = \Lambda\underline{x} + \underline{e} \quad .$$

Upon defining

$$(2.8a)\qquad E(\underline{y}\underline{y}') = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} ,$$

and

$$(2.8b)\qquad E(\underline{x}_1\underline{x}_1') = \Phi \quad ,$$

the structural model may be written as

$$(2.9a)\qquad \Sigma = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} I \\ D \end{bmatrix} \Phi \, [I \;\; D'] + \begin{bmatrix} 0 & 0 \\ 0 & \Theta \end{bmatrix} \cdot \begin{bmatrix} \Lambda_1' & 0 \\ 0 & \Lambda_2' \end{bmatrix} + \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix} ,$$

$$(2.9b)\qquad = \Lambda(B\Phi B' + \Theta)\Lambda' + \Psi \quad ,$$

where

$$\Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \quad ,$$

$$B = \begin{bmatrix} I \\ D \end{bmatrix} \quad ,$$

$$\Phi = \Phi \quad ,$$

$$\Theta = \begin{bmatrix} 0 & 0 \\ 0 & \theta \end{bmatrix} \quad ,$$

and

$$\Psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \quad .$$

The structure given by Equation (2.9b) is more conveniently expressed as

$$\Sigma = \Lambda(B\Phi B' + \Theta)\Lambda' + S_1 * \Psi \quad ,$$

where S_1 is a (2px2p) matrix partitioned into four (pxp) identity matrices. The symbol "$*$" denotes the Hadamard product. The Hadamard operator, in conjunction with the matrix S_1, picks out the diagonal elements of the submatrices of Ψ, which also is partitioned into four (pxp) submatrices. The notation is introduced because it simplifies the expression for the first and second derivatives required for the estimation of parameters.

2.2 Existence and Identification of the Structure

The problem of existence of the structure is essentially the problem of whether the covariance matrix of $\underline{y}$ has the form $\Sigma = \Lambda(B\Phi B + \Theta)\Lambda' + S_1 * \Psi$. In other words, are algebraic solutions for Λ, B, Φ, Θ, and Ψ possible, given the elements of Σ? The problem of identification of the structure, on the other hand, refers to the notion of the uniqueness of Λ, B, Φ, Θ, and Ψ that satisfy the structural model (Anderson and Rubin, 1956). A second way of characterizing the identification problem, following Riersol (1950), is to consider Equation (2.10) to be a structural model that signifies a set of structures. The parameters, given numerically, comprise a structure. A given structure generates one and only one probability distribution $P(\underline{y})$, but more than one structure may generate the same probability distribution. A parameter is identifiable if it has the same value in all such structures. Therefore, a parameter is identifiable if it can be uniquely determined by the probability distribution of the observed variables. Since the probability distribution is a characteristic of the population it should be clear that the identification problem is a problem of the population and would exist even if the population values of the elements of Σ were known.

In principle, one method of determining whether the structure exists is to examine the algebraic conditions necessary to solve the system of equations given by (2.10). However, Anderson and Rubin (1956) point out, in connection with the single occasion common factor model that the algebraic solution is laborious and gives little insignt into the question of the existence of the structure. It can be expected that this situation will also be true for the more complex model considered in this dissertation.

As an alternative approach, a number of authors, (see for example; Anderson and Rubin, 1956; Lawley and Maxwell, 1971; McDonald and Swaminathan, 1972) treat the problems of existence (and identification) through implications of the theory of linear equations. This approach involves the comparison of the number of unknowns and the number of equations. McDonald and Swaminathan (1972) provide the caveat that this approach is of limited value even for linear systems of equations. For a linear system comparison of the rank of the coefficient matrix and the rank of the augmented matrix is the basis for necessary and sufficient conditions for identification. The comparison of the number of unknowns with the number of equations is probably of limited value for non-linear systems of equations, but has been justified by Anderson and Rubin (1956) by the insight it provides into the problem.

The number of independent elements in Σ is $p(2p + 1)$, which correspondingly is the number of equations. There are $2pr$ parameters in Λ, r^2 parameters in D, $\frac{1}{2}r(r + 1)$ parameters in Φ, $\frac{1}{2}r(r + 1)$ parameters in θ and $3p$ parameters in Ψ to be estimated. Thus the total number of unknowns is $2pr + 2r^2 + r + 3p$ while the number of equations is $p(2p + 1)$. However, for any structure, the parameter matrices Λ, D, Φ and θ may be replaced by

$$(2.11a) \qquad \Lambda^* = \begin{bmatrix} \Lambda_1^* & 0 \\ 0 & \Lambda_2^* \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} T_1^{-1} & 0 \\ 0 & T_2^{-1} \end{bmatrix} \quad ,$$

$$(2.11b) \qquad D^* = T_2 D T_1^{-1} \quad ,$$

$$(2.11c) \qquad \Phi^* = T_1 \Phi T_1' \quad ,$$

and

$$(2.11d) \qquad \theta^* = T_2 \theta T_2' \quad ,$$

where $T_1(rxr)$ and $T_2(rxr)$ are non-singular matrices. That is, linear transformations or non-rigid rotations of the matrices are equivalent structures, or alternatively, are admissable structures and will satisfy Equation (2.10). In order to remove this indeterminancy, $2r^2$ restrictions must be placed on the matrices Λ, D, Φ and θ. Taking these restrictions into account, the total number of equations is $p(2p + 1) + 2r^2$ and the difference between the total number of unknowns and the total number of equations is given by

$$(2.12) \qquad \nu = p(2p + 1) - (2pr + 3p + r) \quad .$$

If $\nu \leq 0$, then an algebraic solution is expected to be possible, while if $\nu > 0$, then one can expect a solution only if Σ satisfies ν additional conditions.

The question of identification is essentially the question, if some Λ, D, Φ, θ and Ψ exist that satisfy the model, are these matrices unique? If $\nu < 0$, then it seems $-\nu$ restrictions are necessary for identification. However, if ν is non-negative, then identification can be expected. These considerations lead to the condition

$$\nu = p(2p + 1) - (2pr + 3p + r) \geq 0 \quad ,$$

or

$$r \leq 2p(p - 1)/2p + 1 \quad ,$$

as minimal for identification when the comparison approach is used to examine the identification problem. It will be seen in a later section that the quantity ν is the degress of freedom for the likelihood ratio test that the covariance matrix has the form given by Equation (2.10).

It is evident from Equations (2.11) that the matrices Λ, D, Φ and θ, in general, cannot be identified on the basis of the covariance matrix. If these matrices are to be identifiable they will have to conform to additional restrictions. It is possible to state somewhat more precisely, the conditions under which Λ, D, Φ, and θ will be identified. In fact, it is possible to give sufficient conditions for the identification of Λ. The identification of Λ in turn implies the identification of D, Φ, θ, and ψ_{12}. For the sake of ease of presentation, these conditions are given as theorems. However, as proofs of the theorems are available in other sources, they will not be reproduced in detail here. In some situations a proof will be outlined in order that the conditions may be understood more easily. It should be noted that Theorems 2.1 and 2.5 are slight modifications of theorems reported by Fisher (1966) in his work on the identification problem in econometric models. Theorems 2.3 and 2.4 were originally given as necessary and sufficient conditions for identification of the factor loading matrix for the single occasion common factor analysis model.

Assume that the structure exists and that ψ_{11} and ψ_{22} are identified. Then

(2.13a) $$\Sigma_{11} - \psi_{11} = F_1F_1' \quad ,$$

and

(2.13b) $$\Sigma_{22} - \psi_{22} = F_2F_2' \quad .$$

All equivalent structural characteristics,

(2.14a) $$\Lambda_1 \, \Phi \, \Lambda_1' = F_1F_1' \quad ,$$

and

(2.14b) $$\Lambda_2(D\Phi D' + \theta)\Lambda_2' = F_2F_2' \quad ,$$

must lie in the vector spaces generated by the columns of F_1F_1' and F_2F_2' respectively. Alternatively, all equivalent structural characteristics

(2.15a) $$\Lambda_1 = F_1T_1^{-1} \quad ,$$

and

(2.15b) $$\Lambda_2 = F_2T_2^{-1} \quad ,$$

must lie in the vector spaces spanned by the columns of F_1 and F_2 respectively. The matrices Λ_1, Λ_2, D, Φ and θ are not necessarily the same matrices as those given in Equation (2.11). [The vector spaces generated by the columns of F_1F_1' and F_1 are the same vector spaces. Similarly, the vector spaces generated by the columns of F_2F_2' and F_2 are the same (Rao, 1975)]. Thus, identification of ψ_{11} and ψ_{12} permits the determination of the factor space for Σ_{11} and Σ_{22} but does not permit identification of factor pattern matrices. In order that Λ_1 and Λ_2 be identified up to multiplication on the right by a diagonal matrix, it is necessary and sufficient that the parameters in the two matrices meet restrictions such that the only possible transformation matrices are (rxr) diagonal matrices. A variety of sufficient conditions which imply the identification of Λ_1 and Λ_2, or alternatively, limit all matrices T_1 and T_2 to diagonal matrices are given below. In the first case, a condition that identifies a single column of Λ_1 is given, however, the extension of the condition to the identification of Λ_1 and Λ_2 should be clear.

Assume that linear homogeneous restrictions are placed on the first column of Λ_1. Such restrictions may be expressed as

(2.16) $$R\underline{\lambda}_{11} = \underline{0} \quad .$$

Where R is a (kxp) matrix of constants, and $\underline{\lambda}_{11}$ is the first column of Λ_1. For instance, if the third element of $\underline{\lambda}_{11}$ is zero, a row of R would be

$$r = [0 \; 0 \; 1 \; . \; . \; . \; 0] \quad .$$

If the loadings of the second and fourth variable are equal, then a row of R would be

$$r = [0 \; 1 \; 0 \; -1 \; 0 \; . \; . \; . \; 0] \quad .$$

Denote the first column of T_1^{-1} as $\underline{t}_{11}^{-1}$. In order to determine the conditions for identifiability of $\underline{\lambda}_{11}$ it is necessary to determine the condition under which

(2.17) $$R\Lambda_1\underline{t}_{11}^{-1} = 0 \quad .$$

That is, the condition under which

$$\underline{\lambda}^*_{11} = \Lambda_1 \underline{t}^{-1}_{11} \quad ,$$

satisfies the restrictions

$$R\underline{\lambda}^*_{11} = 0 \quad .$$

Equation (2.17) may be considered to be a system of homogeneous equations in r unknowns which are the elements of $\underline{t}^{-1}_{11}$. The matrix $R\Lambda_1$ has column dimension r. The vector space for the solution of Equation (2.17) will be of dimension one if the rank of $R\Lambda_1$ is r - 1 (Rao, 1975). These considerations lead to the following condition for identification of $\underline{\lambda}_{11}$ up to a scale factor.

<u>Theorem 2.1</u>: A sufficient condition for the identification of $\underline{\lambda}_{11}$ up to a scale factor is that the rank of $R\Lambda_1$ is r - 1.

If the condition of the theorem obtains, then one possible solution for $\underline{t}^{-1}_{11}$ is a vector with a non-zero first element and zero elements elsewhere. Since the dimension of the solution space to Equation (2.17) is one, this vector is the only possible solution vector.

Knowledge of whether the condition expressed in Theorem 2.1 holds appears to require detailed knowledge of Λ_1, which is attainable only if all of the other columns are identifiable. Therefore, the condition may seem of limited usefulness. However, Fisher (1966) distinguishes the two types of cases in which the condition will fail to hold. The first occurs if it is known that another column of Λ_1 obeys the same restrictions. In this case, it is known that the condition fails without complete knowledge of Λ_1. If the first case does not obtain, failure of the condition depends on the last r - 1 columns of $R\Lambda_1$, since the first column is clearly $\underline{0}$ if the restrictions are correct. If a determinant of order r - 1, constructed from the last r - 1 columns of Λ_1 is zero, then the condition fails. However, Fisher (1966) states that such a determinant is a continuous function in the elements of the last r - 1 columns and is linear in any one such column and it, therefore, follows that a vanishing determinant will occur only on a set of measure zero in the space of these elements. Thus, if the first case does not hold, the condition can be assumed to hold in practice.

Suppose that all restrictions on the first column of Λ_1 are such that r - 1 of the $\underline{\lambda}_{i11}$ (i=1,2,...,p) are zero. Such restrictions are referred to as exclusion restrictions. If all restrictions on the columns of, say, Λ_1 are of this form, then the following theorem gives sufficient conditions for identification.

Theorem 2.2: A sufficient condition for the identification of Λ_1 is that the matrix consisting of all rows of Λ_1 which have zeros specified in the mth column (m=1,2,...,r) be rank r - 1 for each value of m (Koopmans and Riersol, 1950).

The conditions of Theorems 2.1 and 2.2 are sufficient rather than necessary and sufficient conditions for identification because restrictions on the parameters in the matrices D, Φ, and θ probably also effect identification of Λ, $\ddot{D}$, Φ, and θ. For instance, suppose that the number of factors is three and the covariance between the first and the third factors is zero. This may be expressed as

$$\underline{t}_{31}\Phi\underline{t}'_{11} = 0 \quad , \tag{2.18}$$

where $\underline{t}_{31}$ is the third row of T_1 and $\underline{t}_{11}$ is the first row. Since $\Phi_{13} = 0$, one way in which Equation (2.18) will be satisfied is if

$$\underline{t}_{31} = [0 \;\; 0 \;\; e_3] \quad , \tag{2.19a}$$

and

$$\underline{t}_{11} = [e_1 \;\; 0 \;\; 0] \quad , \tag{2.19b}$$

where e_1 and e_3 are non-zero. If there are at least r - 2 restrictions on $\underline{\lambda}_{11}$ and $\underline{\lambda}_{31}$ of the form given by Equation (2.16), these solutions given in Equations (2.19) are probably the only possible solutions, but a precise statement of the necessary and sufficient conditions for this to be true cannot be given at present.

Sufficient conditions of a somewhat different nature, for the identification of Λ_1 and Λ_2, may be stated if additional aspects of the model are stipulated. These stipulations and conditions are given in terms of Λ_1, but apply Λ_2 also. One stipulation is that each column of, say, Λ_1 contains at least $\dot{r}$ zeros. Let Λ_{1m} be a submatrix of Λ_1

consisting of the s_m rows that have zeros in the mth column. Further, define Λ^j_{1m} to be a matrix formed by deleting the jth row of Λ_{1m}. The second stipulation is

$$(2.20) \qquad R(\Lambda^j_{1m}) = r - 1 \quad , \qquad m=1,2,\ldots,r; \; j=1,2,\ldots,s_m \quad ,$$

where R( ) indicates the rank of the matrix enclosed in the parentheses. Let Λ_1 be a matrix in which each column has at least r zeros and (2.20) holds, and let

$$C_1 = \Lambda_1 T_1^{-1}$$

be a matrix that does not necessarily have at least r zeros in each column. Further, let

$$(2.21a) \qquad C_{1m} = \Lambda_{1m} T_1^{-1} \quad ,$$

and

$$(2.21b) \qquad C^j_{1m} = \Lambda^j_{1m} T_1^{-1} \quad .$$

It follows from the definitions of Λ_{1m} and Λ^j_{1m} and the rank preserving properties of a non-singular transformation that

$$(2.22a) \qquad R(C_{1m}) = r - 1 \quad ,$$

and

$$(2.22b) \qquad R(C^j_{1m}) = r - 1 \quad .$$

Since A_{1m} contains all of the rows with zeros in the mth column, the addition of a row of A not contained in A_{1m} to A_{1m} will increase the rank of the matrix so formed to r. Therefore, the addition of a row to C_{1m} of a row of C not contained in C_{1m} increases the rank of the matrix so formed to r. Finally, since

$$(2.23) \qquad \Lambda_{1m} = C_{1m}T_1 \quad ,$$

and since the mth column of Λ_{1m} is zero,

$$(2.24) \qquad C_{1m}\underline{t}_{m1} = \underline{0} \quad ,$$

where $\underline{t}_{m1}$ is the mth column of T_1. Since the columns of T_1 are linearly independent, the right nullspace vectors of the matrices C_{1m}, $C_{12},\ldots,C_{1r}$ are linearly independent.

Theorem 2.3: A sufficient condition for identification of Λ_1 up to multiplication on the left by a diagonal matrix is that a matrix $C_1 = \Lambda_1 T_1^{-1}$ contains exactly r submatrices that 1) satisfy (2.22a) and (2.22b), 2) are such that the addition of a row not contained in such a submatrix to such a submatrix increases the rank to r, and 3) have linearly independent right nullspace vectors (Riersol, 1950).

Theorem 2.4: A sufficient condition for the identification of Λ_1 is that Λ_1 does not contain submatrices other than $\Lambda_{11}, \Lambda_{12}, \ldots, \Lambda_{1r}$ that 1) satisfy (2.22a) and (2.22b) with Λ_1 substituted for C_1, and 2) are such that the addition of a row not contained in Λ_{1m} to Λ_{1m} increases its rank to r (Riersol, 1950).

In order to identify Λ_1 and Λ_2 so that even multiplication by a diagonal matrix is not permitted, it is necessary to introduce normalization rules. A normalization rule simply determines the scale for the parameters. Convenient normalization rules consist of specifying one element in each column of Λ_1 and Λ_2 to be unity.

Suppose that identification up to multiplication on the right by a diagonal matrix obtains because the conditions of either Theorem 2.1 or 2.2 are met. Let the focus be on the first column of Λ_1, and let the first element be specified as unity. The normalization may be expressed as the restriction

$$(2.25) \qquad [1\ 0\ .\ .\ .\ 0]\underline{\lambda}_{11} = 1 \quad .$$

The entire set of restrictions on $\underline{\lambda}_{11}$ may be expressed as

$$(2.26) \qquad \begin{bmatrix} 1\ 0\ .\ .\ .\ 0 \\ R \end{bmatrix}\underline{\lambda}_{11} = \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} ,$$

$$S\underline{\lambda}_{11} = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} .$$

Theorem 2.5: A sufficient condition for identification of $\underline{\lambda}_{11}$ is that the rank of $S\Lambda_1$ is r. The rank of $S\Lambda_1$ will be r if and only if the rank of $R\Lambda_1$ is r - 1 (Fisher, 1966).

Suppose that identification up to multiplication on the right by a diagonal matrix obtains because the conditions of either Theorem 2.3 or 2.4 are met. The identification of Λ_1 up to multiplication on the right by a diagonal matrix under these conditions implies that

$$\Lambda_{1m}\underline{t}_{1m}^{-1} = 0 .$$

Since the mth column of Λ_{1m} is 0 and the rank of Λ_{1m} is r - 1, $\underline{t}_{1m}^{-1}$ must be comprised of a non-zero mth element and zero elements elsewhere. Denote the matrix obtained by adding the row with unity in its mth column to A_{1m} as Λ_{1m}^{m}. It follows that

$$(2.27) \qquad \Lambda_{1m}^{m}\underline{t}_{1m}^{-1} = \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} .$$

The matrix Λ_{1m}^{m} is of rank r and, therefore, t_{-1m}^{-1} must now be comprised of a unit mth element and zero elements elsewhere. It should be clear that with the addition of the condition that a normalization is imposed on each column of Λ_1 and Λ_2, Theorems 2.3 and 2.4 give the conditions for identification of Λ_1 and Λ_2.

If Λ_1 and Λ_2 are identified, or alternatively, if identity matrices are the only possible rotation matrices, the matrices D, Φ and θ are also identified. Furthermore, since

$$(2.28) \qquad \Sigma_{12} = \Lambda_1 \Phi D' \Lambda_2' + \psi_{12} \quad ,$$

ψ_{12} must be identified.

The conditions for the identification of the matrices Λ, D, Φ, θ and ψ_{12}, given in Theorems 2.1 to 2.4, are based on the assumption that ψ_{11} and ψ_{22} are identified and, therefore, a critical problem is the identification of the latter two matrices. This is an extremely difficult problem and a solution cannot be offered at present. However, the two generic ways in which identification of these matrices can fail can be indicated. First, suppose that a solution to Equation (2.10) exists for a particular value of r. The possibility arises that there exists more than one ψ_{11} and ψ_{22}, and, therefore, identification of any part of the structure by Theorems 2.1 to 2.4 cannot be achieved. A second possibility that arises is that solutions exist for a different value of r and, again, identification of the structure will not be possible. Thus, the problem of identifying

ψ_{11} and ψ_{22} is fairly complex and more research is required on the problem. Indeed, it must be concluded that more research in general is required on the difficult problems of existence and identification of the structure.

2.3 Restrictions for Identification and Restrictions for Estimation

In chapter three the likelihood equations for estimating the parameters will be given. The likelihood equations cannot be solved in closed form and, therefore, a numerical algorithm for the minimization of a function of several variables is required. A number of the available methods are, therefore, briefly discussed. When applied to the problem of solving the likelihood equations, one similarity among these methods is that the implementation of each procedure requires the restriction of $2r^2$ of the elements of Λ_1, Λ_2, D, Φ and θ. In the discussion of the identification problem, it was mentioned that removing the indeterminancy due to rotation requires the restriction of $2r^2$ of the parameters in Λ_1, Λ_2, D, Φ and θ. Thus both for identification and estimation, $2r^2$ restrictions on the parameters are required.

The reason for this latter requirement is that the likelihood function will attain its maximum everywhere in the r dimensional vector spaces spanned by the columns of Λ_1 and Λ_2, and, as a result, the iterative procedures may break down. However, given that each column of Λ_1 and Λ_2 have been normalized, the solution procedure will not break down provided that restrictions, that meet the conditions of either Theorems 2.1 or 2.2, have been placed on the columns of Λ_1 and Λ_2. When such restrictions are invoked for estimation purposes, an important distinction is between arbitrary and theoretically based restrictions. Theoretically based restrictions usually consist of exclusion restrictions or equality restrictions that are dictated by psychological theory. The restrictions may be indicated by the results of previous factor analyses, by substantiated psychological theory, or by hypotheses about the measurements being investigated. If relevant psychological theory does not exist, arbitrary restrictions that meet the requirements of either Theorem 2.1 or 2.2 can be placed on Λ_1 and Λ_2 and the iterative procedure will not break down.

If arbitrary restrictions are placed on Λ_1 and Λ_2 for estimation purposes, in general, r - 1 restrictions and one normalization will be placed on each column of Λ_1 and Λ_2. These restrictions serve to pick out a representative solution Λ_1, Λ_2, D, Φ and θ from the set of all equivalent solutions $\Lambda_1 T_1^{-1}$, $\Lambda_2 T_2^{-1}$, $T_2' D T_1^{-1}$, $T_1 \Phi T_1$ and $T_2 \theta T_2$. When just r - 1 restrictions and a normalization rule is placed on each column of Λ_1 and Λ_2, the resulting solution is referred to as an unrestricted solution. Any arbitrary unrestricted solution may be obtained from another arbitrary unrestricted solution by the transformation matrices T_1^{-1} and T_2^{-1}. If the r - 1 restrictions are not arbitrary, but reflect psychological theory, then the solution will still be unrestricted. However, it should not be rotated, since rotation would change the values of the loadings that are dictated by psychological theory. However, the solution is obtainable as a rotation of an arbitrary unrestricted solution.

If the restrictions reflect psychological theory, then in some cases, more than r - 1 restrictions that meet the requirements of either Theorem 2.1 or 2.2 may be placed on one or more of the columns of Λ_1 and Λ_2. In this case, the columns Λ_1 and Λ_2 with more than r - 1 restrictions are referred to as overidentified. The solution of the model that incorporates these restrictions is referred to as a restricted solution. In general, a restricted solution cannot be obtained by rotation of an unrestricted solution.

Analyses carried out using arbitrary restrictions have been called exploratory in the factor analysis literature. The purpose of an exploratory analysis is to investigate measurements that are not well understood. Since the longitudinal common factor model permits the study of changes in the parameters of the factor model over occasions, exploratory studies using the model may contribute more to the investigator's understanding of the measurements than a single occasion exploratory factor analysis would. More importantly, for the present discussion, if the exploratory analysis permits the investigator to form hypotheses about the nature of the factors it may be possible to develop hypotheses about the kinds of restrictions the factor loadings conform to. Also, as the nature cf the factors becomes clearer, it may be possible to construct new measurements that are purer measures of the factors. It would seem that if the factors really represent meaningful constructs, it will eventually be possible to construct measures that load on a single factor only. With such measures, the investigator may be able to state restrictions on Λ_1 and Λ_2 that are sufficient to identify Λ, D,

Φ and θ. Moreover, as evidence about restrictions on the factor loadings and other parameters accumulate, the identifiability of ψ_{11} and ψ_{22}, at least for a given value of r, may be effected. Thus, over a series of investigations the identification of the model may be effected.

2.4 Restrictions, Normalizations and Hypothesis Testing

The purpose of this section is to point out that the rejection of some of the hypotheses of interest can be artifacts of the restrictions and normalizations placed on the model. This objective is carried out, in part, by presenting an example in which the normalization rules can be expected to lead to the rejection of the hypothesis of invariant factor patterns. The situation in which restrictions or normalization cause the rejection of an hypothesis should certainly be avoided when the restrictions and normalizations are chosen arbitrarily, as they might be for estimation purposes.

An example of how a normalization rule can effect an hypothesis test is given below. Essentially, a normalization rule sets a scale for the parameters of the model. One possible set of normalization rules is to set the diagonal elements of Φ and $D\Phi D' + \theta$ to unity. Unfortunately, these normalization rules can lead to problems in testing the hypothesis $\Lambda_1 = \Lambda_2$. In general, the variance of the observed variables may be expected to increase over occasions. Lord (1963) has shown that the variance of a variable will increase unless either of two conditions are fulfilled. For any variable z, let z_1 indicate the variable at the first occasion and z_2 the variable at the second occasion.

If

$$G = z_2 - z_1 \quad ,$$

then

$$\sigma^2_{z_2} = \sigma^2_G + \sigma^2_{z_1} + 2\sigma_{Gz_1} \quad .$$

The variance of z will increase unless either σ^2_G is zero or there is a negative covariance σ_{Gz_1} such that

$$\text{(2.29)} \qquad \sigma_{Gz_1} \leq -\sigma^2_G/2 \quad .$$

The variance will remain the same only if σ^2_G is zero or Equation (2.32) is an equality. Since each of these conditions is fairly restrictive it seems likely that the variance of at least one variable will increase.

If the variance of at least one of the variables increases and the diagonal elements of Φ and $D\Phi D + \theta$ are required to be unity, then the elements of Λ_2 corresponding to that variable must increase relative to the corresponding elements in Λ_1. However, since one of the purposes of the model is to examine the question of the equality of factor loadings this situation should not be permitted. In order not to have the normalization rules lead to rejection of the hypothesis of equal factor patterns, normalization rules can be introduced that specify one entry in each column of Λ_1 and Λ_2 to be unity, as was suggested in section 2.2.

The dilemma that now arises is that if both sets of normalizations are arbitrary, why should the results based on one set be chosen over the results based on another? The answer is that the normalization rules that require the diagonal elements of Φ and $D\Phi D + \theta$ to be unity, although arbitrarily chosen, in fact are quite restrictive. These normalization rules imply that the factor variances remain the same over the two occasions. But this can happen only if the factor scores meet either of the conditions given above. The question of whether the first condition obtains can be tested statistically. This is the question of factor score invariance. The second condition cannot be tested statistically, but it seems unlikely that it would hold simultaneously for all factors. Therefore, the normalization rule that sets certain elements of Λ_1 and Λ_2 to be unity seems to be more viable.

In a similar vein, if the equality of factor pattern matrices is being entertained, the same restrictions and normalizations must be placed on corresponding elements of the columns of Λ_1 and Λ_2. This, of course, means that some of the factor loadings will be equal by fiat and, therefore, it is clear that the restrictions and normalization must limit the hypotheses that can be tested. It is important to be aware of these limitations and to whatever extent possible choose normalizations and restrictions that permit testing of hypotheses that are coincident with the purposes of the study. However, again, restrictions and normalization should not be manipulated to conform with hypothesis testing purposes, if the manipulations will produce results that are at odds with the results expected on theoretical grounds.

CHAPTER III

ESTIMATION OF PARAMETERS

3.0 Introduction

In this chapter the likelihood equations for estimating the parameters of the model are given. These results are presented in section 3.1. Methods for the solution of the likelihood equations are discussed in section 3.2 and the second derivatives required for certain of the procedures are presented. In section 3.3, issues related to changing the units of measurement of the observed variables are discussed.

3.1 The Maximum Likelihood Equations

In order to justify the use of a maximum likelihood estimation procedure, the random vectors $\underline{x}$ and $\underline{e}$ formed in (2.7) are assumed to follow independent multivariate normal distributions. The random vector $\underline{y}$ is, of course, assumed to have the form given by (2.7) and so also follows a multivariate normal distribution. The assumed covariance structure of $\underline{y}$ is given by Equation (2.10).

Suppose a random sample of N observations $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_N$ is available. The unbiased sample estimate S of the population covariance matrix is given by

$$(3.1) \qquad S = \frac{1}{N-1} \sum_{i=1}^{N} (\underline{y}_i - \underline{y})(\underline{y}_i - \underline{y})' \quad ,$$

where $\bar{\underline{y}}$ is the sample mean vector. The problem is to use the information contained in S to estimate the parameter matrices Λ, B, Φ, Θ and Ψ of the model. This may be done by finding the values of the parameters that maximize the likelihood function of the observations $\underline{y}_i$. Since the model given by Equation (2.10) does not depend in any way on the mean vector it is sufficient to maximize the natural logarithm of the likelihood function based on only the information contained in S. This function, omitting a function of the observations, is given by

$$\text{(3.2)} \qquad \log L = -\frac{n}{2}[\log|\Sigma| + \mathrm{Tr}S\Sigma^{-1}] \quad ,$$

where

$$n = N - 1 \quad .$$

However, Joreskog (1970) has suggested that it is more convenient to minimize the function,

$$\text{(3.3)} \qquad F(\Lambda,B,\Phi,\Theta,\Psi) = \log|\Sigma| + \mathrm{Tr}(S\Sigma^{-1}) - \log|S| - 2p \quad .$$

Minimization of (3.3) is equivalent to a maximization of (3.2) and so the function given by (3.3) will be used in the sequel. In order to obtain the first derivatives of (3.3) the matrix calculus developed by McDonald and Swaminathan (1973) will be used. For the reader's convenience, the results that are required to obtain the derivatives are given in Appendix A. The derivatives required to minimize (3.3) are $\delta F/\delta\Lambda_{diag}$, $\delta F/\delta D$, $\delta F/\delta\Phi$, $\delta F/\delta\theta$ and $\delta F/\delta\Psi$, where

$$\Lambda_{diag} = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix} .$$

It may be shown that

$$(3.4) \qquad \frac{\delta F}{\delta\Sigma} = \frac{\delta[\log|\Sigma| + Tr(S\Sigma^{-1})]}{\delta\Sigma} ,$$

$$= Vec(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1}) .$$

Now by the chain rule (Theorem A.1),

$$\frac{\delta\Sigma}{\delta\Lambda_{diag}} = \frac{\delta\Lambda}{\delta\Lambda_{diag}} \frac{\delta\Sigma}{\delta\Lambda} .$$

With the use of Theorem A.3,

$$(3.5) \qquad \frac{\delta\Sigma}{\delta\Lambda} = \frac{\delta\Lambda}{\delta\Lambda} [I \otimes\otimes (B\Phi B' + \Theta)\Lambda'] + \frac{\delta\Lambda'}{\delta\Lambda} [(B\Phi B' + \Theta)\Lambda' \otimes\otimes I] .$$

The symbol '$\otimes\otimes$' indicates the Double Kronecker product, which is defined by Equation (A.16). By the chain rule

$$\frac{\delta F}{\delta\Lambda_{diag}} = \frac{\delta\Lambda}{\delta\Lambda_{diag}} \frac{\delta\Sigma}{\delta\Lambda} \frac{\delta F}{\delta\Sigma} .$$

On letting

$$W = \Sigma^{-1} S \Sigma^{-1} \quad ,$$

and

$$\Gamma = B\Phi B' + \Theta \quad ,$$

we have

$$\frac{\delta F}{\delta \Lambda_{diag}} = \frac{\delta \Lambda}{\delta \Lambda_{diag}} \left[\frac{\delta \Lambda}{\delta \Lambda} (I \otimes\otimes \Gamma\Lambda') + \frac{\delta \Lambda'}{\delta \Lambda} (\Gamma\Lambda' \otimes\otimes I) \right] [\mathrm{Vec}(\Sigma^{-1} - W)] \quad .$$

Using Theorem A.7 we obtain

$$\frac{\delta F}{\delta \Lambda_{diag}} = \frac{\delta \Lambda}{\delta \Lambda_{diag}} \left[\mathrm{Vec}(\Sigma^{-1} - W)\Lambda\Gamma + \frac{\delta \Lambda'}{\delta \Lambda} \mathrm{Vec}\Gamma\Lambda'(\Sigma^{-1} - W) \right] \quad .$$

Application of Theorem A.4 yields

$$\text{(3.6)} \qquad \frac{\delta F}{\delta \Lambda_{diag}} = \frac{\delta \Lambda}{\delta \Lambda_{diag}} [\mathrm{Vec}(\Sigma^{-1} - W)\Lambda\Gamma + \mathrm{Vec}(\Sigma^{-1} - W)\Lambda\Gamma] \quad ,$$

$$= 2\mathrm{Vec}[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag} \quad ,$$

where $\mathrm{Vec}(A)_{diag}$ is defined by (A.15).

Again by the chain rule,

$$\frac{\delta F}{\delta D} = \frac{\delta B}{\delta D}\frac{\delta \Sigma}{\delta B}\frac{\delta F}{\delta \Sigma} \quad .$$

Since

(3.7) $$\frac{\delta \Sigma}{\delta B} = (\Lambda' \otimes \otimes \Phi B'\Lambda') + \frac{\delta B'}{\delta B}(\Phi B'\Lambda' \otimes \otimes \Lambda') \quad ,$$

we have

$$\frac{\delta F}{\delta D} = \frac{\delta B}{\delta D}(\Lambda' \otimes \otimes B'\Lambda')\, \mathrm{Vec}(\Sigma^{-1} - W)$$

$$+ \frac{\delta B'}{\delta B}[(\Phi B'\Lambda' \otimes \otimes \Lambda')\, \mathrm{Vec}(\Sigma^{-1} - W)] \quad ,$$

which, using Theorems A.4 and A.7, simplifies to

(3.8) $$\frac{\delta F}{\delta D} = \frac{\delta B}{\delta D}[2\mathrm{Vec}\Lambda'(\Sigma^{-1} - W)\Lambda B\Phi] \quad ,$$

where

$$\frac{\delta B}{\delta D} = [\,0 \quad I_{r^2}] \quad .$$

Hence,

$$(3.9) \qquad \frac{\delta F}{\delta D} = \Lambda_2'(\Sigma^{21} - W_{21})\Lambda_1\Phi + \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 D\Phi \quad ,$$

where

$$\Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \quad ,$$

and

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} \quad .$$

Since, by Theorem A.2

$$(3.10) \qquad \frac{\delta \Sigma}{\delta \Phi} = \frac{\delta \Phi}{\delta \Phi} (B'\Lambda' \otimes\otimes B'\Lambda') \quad ,$$

and by the chain rule

$$\frac{\delta F}{\delta \Phi} = \frac{\delta \Phi}{\delta \Phi} \frac{\delta \Sigma}{\delta \Phi} \frac{\delta F}{\delta \Sigma} \quad ,$$

we obtain

$$\frac{\delta F}{\delta \Phi} = \frac{\delta \Phi}{\delta \Phi} (B'\Lambda' \otimes\otimes B'\Lambda') \operatorname{Vec}(\Sigma^{-1} - W) \quad ,$$

$$= \frac{\delta \Phi}{\delta \Phi} \operatorname{Vec}[B'\Lambda'(\Sigma - W)\Lambda B] \quad .$$

By Theorem A.6 and symmetry considerations

$$(3.11)\qquad \frac{\delta F}{\delta \Phi} = \mathrm{Vec}[2B'\Lambda'(\Sigma^{-1} - W)\Lambda B - I * B'\Lambda'(\Sigma^{-1} - W)\Lambda B] \quad .$$

Further,

$$\frac{\delta F}{\delta \theta} = \frac{\delta \Theta}{\delta \theta}\,\frac{\delta \Sigma}{\delta \Theta}\,\frac{\delta F}{\delta \Sigma} \quad .$$

Since,

$$(3.12)\qquad \frac{\delta \Sigma}{\delta \Theta} = \frac{\delta \Theta}{\delta \Theta}\,(\Lambda' \otimes \otimes \Lambda') \quad ,$$

it follows that

$$(3.13)\qquad \frac{\delta F}{\delta \theta} = \frac{\delta \Theta}{\delta \theta}\;\frac{\delta \Theta}{\delta \Theta}\,(\Lambda' \otimes \otimes \Lambda')\ \mathrm{Vec}(\Sigma^{-1} - W) \quad ,$$

$$= \frac{\delta \Theta}{\delta \theta}\;\frac{\delta \Theta}{\delta \Theta}\,\mathrm{Vec}\Lambda'(\Sigma^{-1} - W)\Lambda \quad ,$$

$$= \frac{\delta \Theta}{\delta \theta}\,\mathrm{Vec}\Lambda'(\Sigma^{-1} - W)\Lambda \quad ,$$

where

$$(3.14)\qquad \frac{\delta \Theta}{\delta \theta} = [\ 0 \quad 0 \quad 0 \quad L_{r^2}\ ] \quad .$$

The matrix L_{r^2} $(r^2 x r^2)$ is defined by (A.2). Thus

(3.15) $$\frac{\delta F}{\delta \theta} = \mathrm{Vec}[2\Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 - I * \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2] \quad .$$

Since

$$\frac{\delta F}{\delta \Psi} = \frac{\delta \Sigma}{\delta \Psi}\frac{\delta F}{\delta \Psi} \quad ,$$

and since in obtaining the partial derivative $\delta F/\delta\Psi$, Σ is linear in Ψ, the derivative becomes

$$\frac{\delta F}{\delta \Psi} = \frac{\delta(S_1 * \Psi)}{\delta \Psi} \; \frac{\delta F}{\delta \Sigma} \quad ,$$

(3.16) $$= D(S_1)L^* \, \mathrm{Vec}(\Sigma^{-1} - W) \quad ,$$

where

$$L^* = \frac{\delta \Psi}{\delta \Psi} = \begin{bmatrix} L_{p^2} & 0 & 0 & 0 \\ 0 & I_{p^2} & E_{p^2} & 0 \\ 0 & E_{p^2} & I_{p^2} & 0 \\ 0 & 0 & 0 & L_{p^2} \end{bmatrix} \quad .$$

The matrix E_{p^2} $(p^2 x p^2)$ is defined by (A.4). Using Theorem A.6

$$\frac{\delta F}{\delta \Psi} = D(S_1)\text{Vec}[2(\Sigma^{-1} - W) - I * (\Sigma^{-1} - W)] \quad .$$

The matrix $D(S_1)$ is a $(4p^2 x 4p^2)$ supermatrix partitioned into sixteen $(p^2 x p^2)$ submatrices. Each of the four principal diagonal submatrices of $D(S_1)$ has unity as its (g,g)th diagonal element if $g = p(j - 1) + j$, $(j=1,2,\ldots,p)$ and zeros elsewhere. For example, let $p = 5$, then in each of the principal diagonal submatrices in $D(S_1)$, the first, seventh, thirteenth, nineteenth and twenty-fifth diagonal elements are unity. All other elements in the principal diagonal submatrices are zero, as are all the elements of the off diagonal submatrices. The non-zero elements of $\delta F/\delta \Psi$ may be expressed as

(3.17a) $$\frac{\delta F}{\delta \psi_{11\ \text{diag}}} = (\Sigma^{11} - W_{11})_{\text{diag}} \quad ,$$

(3.17b) $$\frac{\delta F}{\delta \psi_{12\ \text{diag}}} = \frac{\delta F}{\delta \psi_{21\ \text{diag}}} = 2(\Sigma^{12} - W_{12})_{\text{diag}} \quad ,$$

and

(3.17c) $$\frac{\delta F}{\delta \psi_{22\ \text{diag}}} = (\Sigma^{22} - W_{22})_{\text{diag}} \quad ,$$

where, say, $(\Sigma^{11} - W_{11})_{\text{diag}}$ is a (px1) vector consisting of the diagonal elements of $(\Sigma^{11} - W_{11})$.

Thus, the likelihood equations that must be solved to minimize F are

$$2[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag} = 0 \quad , \tag{3.18a}$$

$$\Lambda_2'(\Sigma^{21} - W_{21})\Lambda_1\Phi + \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 D\Phi = 0 \quad , \tag{3.18b}$$

$$2B'\Lambda'(\Sigma^{-1} - W)\Lambda B - I * B'\Lambda'(\Sigma^{-1} - W)\Lambda B = 0 \quad , \tag{3.18c}$$

$$2\Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 - I * \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 = 0 \quad , \tag{3.18d}$$

$$(\Sigma^{11} - W_{11})_{diag} = 0 \quad , \tag{3.18e}$$

$$2(\Sigma^{12} - W_{12})_{diag} = 0 \quad , \tag{3.18f}$$

and

$$(\Sigma^{22} - W_{22})_{diag} = 0 \quad . \tag{3.18g}$$

3.2 Solution of the Likelihood Equation

Equations (3.18) apparently cannot be solved in closed form and, therefore, an iterative method is required for a solution. The available mathematically justified methods for the unrestrained minimization of a function of several variables all use the following equation:

$$\underline{\Theta}^{(k+1)} = \underline{\Theta}^{(k)} - \lambda^{(k)}[M(\underline{\Theta}^{(k)})]^{-1}f'(\underline{\Theta}^{(k)}) \quad ,$$

where k is the iteration index. The procedures differ in the definition of $\lambda^{(k)}$ and $M(\underline{\Theta}^{(k)})$ and may be thought of as falling into two classes based on the definition of $M(\underline{\Theta}^{(k)})$. The first class, comprised of the Newton-Raphson procedure and modified Newton-Raphson procedures involve the Hessian,

$$H(\underline{\Theta}) = \begin{bmatrix} \frac{\delta^2 F}{\delta\Lambda_{diag}\delta\Lambda_{diag}} & & & & \\ \frac{\delta^2 F}{\delta D\delta\Lambda_{diag}} & \frac{\delta^2 F}{\delta D\delta D} & & & \\ \frac{\delta^2 F}{\delta\Phi\delta\Lambda_{diag}} & \frac{\delta^2 F}{\delta\Phi\delta D} & \frac{\delta^2 F}{\delta\Phi\delta\Phi} & & \\ \frac{\delta^2 F}{\delta\theta\delta\Lambda_{diag}} & \frac{\delta^2 F}{\delta\theta\delta D} & \frac{\delta^2 F}{\delta\theta\delta\Phi} & \frac{\delta^2 F}{\delta\theta\delta\theta} & \\ \frac{\delta^2 F}{\delta\Psi\delta\Lambda_{diag}} & \frac{\delta^2 F}{\delta\Psi\delta D} & \frac{\delta^2 F}{\delta\Psi\delta\Phi} & \frac{\delta^2 F}{\delta\Psi\delta\theta} & \frac{\delta^2 F}{\delta\Psi\delta\Psi} \end{bmatrix}$$

where $\underline{\Theta}$ is a vector of dimension $(2pr + 3p + r)$ at most. The vector $\underline{\Theta}^{(k)}$ contains the estimates of the free parameters of Λ, B, Φ, Θ and Ψ given by the kth iteration. The Hessian is evaluated using these estimates and substituted for $M(\underline{\Theta}^{(k)})$, in the Newton-Raphson and modified Newton-Raphson procedures. These two procedures, therefore, require expressions for the second derivatives. The second class of procedures vary in their definition of $M(\underline{\Theta}^{(k)})$, but do not require expressions for the second derivatives. The most well known methods in this class are The Fletcher-Powell Method and the classical Gauss-Seidel procedure. In the remainder of this section expressions for the second derivatives of the likelihood function are derived. These expressions, provide the potential for utilizing the Newton-Raphson or modified Newton-Raphson procedures for the solution of the likelihood equations. In addition, the second derivatives provide the information necessary for obtaining the asymptotic standard errors of the estimates.

In order to simplify the derivations it is useful to evaluate a few basic derivatives. It can be shown that

$$(3.19)\qquad \frac{\delta(\Sigma^{-1} - W)}{\delta\Sigma} = \Sigma^{-1} \otimes\!\otimes (W - \Sigma^{-1}) + (W \otimes\!\otimes \Sigma^{-1}) \quad .$$

Now,

$$\frac{\delta(\Sigma^{-1} - W)}{\delta\Lambda} = \frac{\delta\Sigma}{\delta\Lambda}\ \frac{\delta(\Sigma^{-1} - W)}{\delta\Sigma} \quad .$$

By Equations (3.6) and (3.19)

$$(3.20)\qquad \frac{\delta(\Sigma^{-1} - W)}{\delta\Lambda} = [\frac{\delta\Lambda}{\delta\Lambda}\ (I \otimes\!\otimes \Gamma\Lambda') + \frac{\delta\Lambda'}{\delta\Lambda}\ (\Gamma\Lambda' \otimes\!\otimes I)]$$

$$[\Sigma^{-1} \otimes\!\otimes (W - \Sigma^{-1}) + W \otimes\!\otimes \Sigma^{-1}] \quad ,$$

$$= \frac{\delta\Lambda}{\delta\Lambda} [\Sigma^{-1} \otimes\otimes \Gamma\Lambda'(W - \Sigma^{-1}) + W \otimes\otimes \Gamma\Lambda'\Sigma^{-1}]$$

$$+ \frac{\delta\Lambda'}{\delta\Lambda} [\Gamma\Lambda'\Sigma^{-1} \otimes\otimes (W - \Sigma^{-1}) + \Gamma\Lambda' \otimes\otimes \Sigma^{-1}] \quad .$$

Similarly,

$$\frac{\delta(\Sigma^{-1} - W)}{\delta B} = \frac{\delta\Sigma}{\delta B} \frac{\delta(\Sigma^{-1} - W)}{\delta\Sigma} \quad ,$$

and by Equations (3.7) and (3.19)

$$\frac{\delta(\Sigma^{-1} - W)}{\delta B} = [\frac{\delta B}{\delta B} (\Lambda' \otimes\otimes \Phi B'\Lambda') + \frac{\delta B'}{\delta B} (\Phi B'\Lambda' \otimes\otimes \Lambda')]$$

$$[\Sigma^{-1} \otimes\otimes (W - \Sigma^{-1}) + W \otimes\otimes \Sigma^{-1}] \quad ,$$

which reduces to

$$\frac{\delta(\Sigma^{-1} - W)}{\delta B} = \frac{\delta B}{\delta B} [\Lambda'\Sigma^{-1} \otimes\otimes \Phi B'\Lambda'(W - \Sigma^{-1}) + \Lambda'W \otimes\otimes \Phi B'\Lambda'\Sigma^{-1}] \tag{3.21}$$

$$+ \frac{\delta B'}{\delta B} [\Phi B'\Lambda'\Sigma^{-1} \otimes\otimes \Lambda'(W - \Sigma^{-1}) + \Phi B'\Lambda'W \otimes\otimes \Lambda'\Sigma^{-1}] \quad .$$

Using almost identical steps,

$$\frac{\delta(\Sigma^{-1} - W)}{\delta\Phi} = \frac{\delta\Phi}{\delta\Phi} [B'\Lambda'\Sigma^{-1} \otimes\otimes B'\Lambda'(W - \Sigma^{-1}) + B'\Lambda'W \otimes\otimes B'\Lambda'\Sigma^{-1}] \quad , \tag{3.22}$$

$$= L_{r^2}[B'\Lambda'\Sigma^{-1} \otimes\otimes B'\Lambda'(W - \Sigma^{-1}) + B'\Lambda'W \otimes\otimes B'\Lambda'\Sigma^{-1}] \quad .$$

Also,

(3.23) $$\frac{\delta(\Sigma^{-1} - W)}{\delta\Theta} = \frac{\delta\Theta}{\delta\Theta}[\Lambda'\Sigma^{-1} \otimes\otimes \Lambda'(W - \Sigma^{-1}) + \Lambda'W \otimes\otimes \Lambda'\Sigma^{-1}] \quad .$$

Finally,

(3.24) $$\frac{\delta(\Sigma^{-1} - W)}{\delta\Psi} = D(S_1)L^*[\Sigma^{-1} \otimes\otimes (W - \Sigma^{-1}) + W \otimes\otimes \Sigma^{-1}] \quad .$$

Now,

$$\frac{\delta F}{\delta\Lambda} = 2\text{Vec}(\Sigma^{-1} - W)\Lambda\Gamma \ .$$

Using Equation (3.20)

$$\frac{\delta^2 F}{\delta\Lambda\delta\Lambda} = 2\{\frac{\delta(\Sigma^{-1} - W)}{\delta\Lambda}(I \otimes\otimes \Lambda\Gamma) + \frac{\delta\Lambda\Gamma}{\delta\Lambda}[(\Sigma^{-1} - W) \otimes\otimes I]\} \quad ,$$

$$= 2\{\frac{\delta\Lambda}{\delta\Lambda}[\Sigma^{-1} \otimes\otimes \Gamma\Lambda'(W - \Sigma^{-1})\Lambda\Gamma + W \otimes\otimes \Gamma\Lambda'\Sigma^{-1}\Lambda\Gamma]$$

$$+ \frac{\delta\Lambda'}{\delta\Lambda}[\Gamma\Lambda'\Sigma^{-1} \otimes\otimes (W - \Sigma^{-1})\Lambda\Gamma + \Gamma\Lambda'W \otimes\otimes \Sigma^{-1}\Lambda\Gamma]$$

$$+ \frac{\delta\Lambda}{\delta\Lambda}[(\Sigma^{-1} - W) \otimes\otimes \Gamma]\} \quad .$$

However, the required second derivative is $\delta^2 F/\delta\Lambda_{diag}\delta\Lambda_{diag}$.
Utilizing Theorem A.4, we obtain

$$(3.25)\qquad \frac{\delta^2 F}{\delta\Lambda_{diag}\delta\Lambda_{diag}} = [\Sigma^{-1}\circledast\Gamma\Lambda'(W - \Sigma^{-1})\Lambda\Gamma + W\circledast\Gamma\Lambda'\Sigma^{-1}\Lambda\Gamma]$$

$$+ E_1[\Gamma\Lambda'\Sigma^{-1}\circledast(W - \Sigma^{-1})\Lambda\Gamma + \Gamma\Lambda'W\circledast\Sigma^{-1}\Lambda\Gamma]$$

$$+ [(\Sigma^{-1} - W)\circledast\Gamma] \quad ,$$

where

$$E_1 = \frac{\delta\Lambda'_{diag}}{\delta\Lambda_{diag}} = \begin{bmatrix} E_{pr} & 0 \\ 0 & E_{pr} \end{bmatrix} \quad ,$$

and the symbol ' ' denotes the Hadamard-Kronecker product defined by (A.22). In order to obtain an expression for

$$\frac{\delta^2 F}{\delta D\delta\Lambda_{diag}} = \frac{\delta B}{\delta D}\,\frac{\delta^2 F}{\delta B\delta\Lambda}\,\frac{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag}}{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]} \quad ,$$

we first obtain the second partial derivative

$$\frac{\delta^2 F}{\delta B\delta\Lambda} = 2\{\frac{\delta(\Sigma^{-1} - W)}{\delta B}\}[I\otimes\otimes\Lambda\Gamma] + \frac{\delta\Lambda\Gamma}{\delta B}\,[(\Sigma^{-1} - W)\otimes\otimes I] \quad ,$$

$$= 2\,\{\frac{\delta B}{\delta B}\,[\Lambda'\Sigma^{-1} \otimes\!\!\otimes\, \Phi B'\Lambda'(W - \Sigma^{-1})\Lambda\Gamma + \Lambda' W \otimes\!\!\otimes\, \Phi B'\Lambda'\Sigma^{-1}\Lambda\Gamma$$

$$+ \Lambda'(\Sigma^{-1} - W) \otimes\!\!\otimes\, \Phi B] + \frac{\delta B'}{\delta B}\,[\Phi B'\Lambda'\Sigma^{-1} \otimes\!\!\otimes\, \Lambda'(W - \Sigma^{-1})\Lambda\Gamma$$

$$+ \Phi B'\Lambda' W \otimes\!\!\otimes\, \Lambda'\Sigma^{-1}\Lambda\Gamma + \Phi B'\Lambda'(\Sigma^{-1} - W) \otimes\!\!\otimes\, I]\} \quad .$$

On letting

$$(3.26a) \qquad P = \frac{\delta B}{\delta D} = [\; 0 \quad I_{r^2} \;] \quad ,$$

and

$$(3.26b) \qquad Q = \frac{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag}}{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & I_{pr} \end{bmatrix} \quad ,$$

we have

$$(3.27) \qquad \frac{\delta^2 F}{\delta D \delta \Lambda_{diag}} = 2P[\Lambda'\Sigma^{-1} \otimes\!\!\otimes\, \Phi B'\Lambda'(W - \Sigma^{-1})\Lambda\Gamma + \Lambda' W \otimes\!\!\otimes\, \Phi B'\Lambda'\Sigma^{-1}\Lambda\Gamma$$

$$+ \Lambda'(\Sigma^{-1} - W) \otimes\!\!\otimes\, \Phi B' + \Phi B'\Lambda'\Sigma^{-1} \otimes\!\!\otimes\, \Lambda'(W - \Sigma^{-1})\Lambda\Gamma$$

$$+ \Phi B'\Lambda' W \otimes\!\!\otimes\, \Lambda'\Sigma^{-1}\Lambda\Gamma + \Phi B'\Lambda'(\Sigma^{-1} - W) \otimes\!\!\otimes\, I]Q \quad .$$

Similarly, to obtain an expression for

$$\frac{\delta^2 F}{\delta\Phi\delta\Lambda_{diag}} = \frac{\delta^2 F}{\delta\Phi\delta\Lambda} \frac{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag}}{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]} ,$$

we first evaluate the second partial derivative

$$\frac{\delta^2 F}{\delta\Phi\delta\Lambda} = 2\left\{\frac{\delta(\Sigma^{-1} - W)}{\delta\Phi}(I \otimes\otimes \Lambda\Gamma) + \frac{\delta\Lambda\Gamma}{\delta\Phi}[(\Sigma^{-1} - W)\otimes\otimes I]\right\} ,$$

$$= 2L_{r^2}[B'\Lambda'\Sigma^{-1} \otimes\otimes B'\Lambda'(W - \Sigma^{-1})\Lambda\Gamma$$

$$+ B'\Lambda'W \otimes\otimes B'\Lambda'\Sigma^{-1}\Lambda\Gamma + B'\Lambda'(\Sigma^{-1} - W)\otimes\otimes B'] .$$

Then by Equation (3.26b)

$$\frac{\delta^2 F}{\delta\Phi\delta\Lambda_{diag}} = 2L_{r^2}[B'\Lambda'\Sigma^{-1} \otimes\otimes B'\Lambda'(W - \Sigma^{-1})\Lambda\Gamma$$
$$+ B'\Lambda'W \otimes\otimes B'\Lambda'\Sigma^{-1}\Lambda\Gamma + B'\Lambda'(\Sigma^{-1} - W)\otimes\otimes B']Q . \tag{3.28}$$

Now,

$$\frac{\delta^2 F}{\delta\theta\delta\Lambda_{diag}} = \frac{\delta\Theta}{\delta\theta} \frac{\delta^2 F}{\delta\Theta\delta\Lambda} \frac{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]_{diag}}{\delta[(\Sigma^{-1} - W)\Lambda\Gamma]} ,$$

and

$$\frac{\delta^2 F}{\delta\Theta\delta\Lambda} = 2\left\{\frac{\delta(\Sigma^{-1} - W)}{\delta\Theta}(I \otimes\otimes \Lambda\Gamma) + \frac{\delta\Lambda\Gamma}{\delta\Theta}[(\Sigma^{-1} - W)\otimes\otimes I]\right\} ,$$

$$= 2\,\frac{\delta\Theta}{\delta\Theta}[\Lambda'\Sigma^{-1}\otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda\Gamma + \Lambda'W\otimes\otimes\Lambda'\Sigma^{-1}\Lambda\Gamma$$

$$+ \Lambda'(\Sigma^{-1} - W)\otimes\otimes I] .$$

By Equation (3.14)

$$R = \frac{\delta\Theta}{\delta\theta} = [\ 0 \quad 0 \quad 0 \quad L_{r^2}\ ] .$$

Hence, using Equation (3.26b), we obtain

$$\frac{\delta^2 F}{\delta\theta\delta\Lambda_{diag}} = 2R[\Lambda'\Sigma^{-1}\otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda\Gamma + \Lambda'W\otimes\otimes\Lambda'\Sigma^{-1}\Lambda\Gamma \tag{3.29}$$

$$+ \Lambda'(\Sigma^{-1} - W)\otimes\otimes I]Q .$$

Further,

$$\frac{\delta^2 F}{\delta\Psi\delta\Lambda} = 2\,\frac{\delta(\Sigma^{-1} - W)}{\delta\Psi}(I\otimes\otimes\Lambda\Gamma) ,$$

$$= 2D(S_1)L^*[\Sigma^{-1}\otimes\otimes (W - \Sigma^{-1})\Lambda\Gamma + W\otimes\otimes\Sigma^{-1}\Lambda\Gamma] .$$

Hence,

$$\frac{\delta^2 F}{\delta\Psi\delta\Lambda_{diag}} = 2D(S_1)L^*[\Sigma^{-1}\otimes\otimes (W - \Sigma^{-1})\Lambda\Gamma + W\otimes\otimes\Sigma^{-1}\Lambda\Gamma]Q . \tag{3.30}$$

Now, rewriting Equation (3.8), we have

$$\frac{\delta F}{\delta D} = \frac{\delta B}{\delta D} 2\mathrm{Vec}\Lambda'(\Sigma^{-1} - W)\Lambda B\Phi \quad ,$$

where again

$$P = \frac{\delta B}{\delta D} [\ 0 \quad I_{r^2}\ ] \quad .$$

This derivative may also be expressed as

$$\frac{\delta F}{\delta D} = P^*\Lambda'(\Sigma^{-1} - W)\Lambda B\Phi \quad ,$$

where P* (rx2r) is

$$P^* = [\ 0 \quad I_r\ ] \quad .$$

By the chain rule

$$\frac{\delta^2 F}{\delta D\delta D} = \frac{\delta B}{\delta D}\frac{\delta^2 F}{\delta B\delta D} \quad .$$

Now

$$\frac{\delta^2 F}{\delta B\delta D} = 2 \quad \frac{\delta[P^*\Lambda'(\Sigma^{-1} - W)\Lambda]}{\delta B} (I \otimes\otimes B\Phi)$$

$$+ \frac{\delta B\Phi}{\delta B} [\Lambda'(\Sigma^{-1} - W)\Lambda P^{*\prime} \otimes\otimes I] \quad ,$$

which by virtue of Equation (3.21) becomes

$$\frac{\delta^2 F}{\delta B \delta D} = 2\{\frac{\delta B}{\delta B}[\Lambda'\Sigma^{-1}\Lambda P^{*'} \otimes\otimes \Phi B'\Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ \Lambda' W \Lambda P^{*'} \otimes\otimes \Phi B'\Lambda'\Sigma^{-1}\Lambda B\Phi + \Lambda'(\Sigma^{-1} - W)\Lambda P^{*'} \otimes\otimes \Phi]$$

$$+ \frac{\delta B'}{\delta B}[\Phi B'\Lambda'\Sigma^{-1}\Lambda P^{*'} \otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ \Phi B'\Lambda' W \Lambda P^{*'} \otimes\otimes \Lambda'\Sigma^{-1}\Lambda B\Phi]\} \quad .$$

Hence,

(3.31) $$\frac{\delta^2 F}{\delta D \delta D} = 2P[\Lambda'\Sigma^{-1}\Lambda P^{*'} \otimes\otimes \Phi B'\Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ \Lambda' W \Lambda P^{*'} \otimes\otimes \Phi B'\Lambda'\Sigma^{-1}\Lambda B\Phi + \Lambda'(\Sigma^{-1} - W)\Lambda P^{*'} \otimes\otimes \Phi$$

$$+ \Phi B'\Lambda'\Sigma^{-1}\Lambda P^{*'} \otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ \Phi B'\Lambda' W \Lambda P^{*'} \otimes\otimes \Lambda'\Sigma^{-1}\Lambda B\Phi] \quad .$$

Similarly, using (3.22)

(3.32) $$\frac{\delta^2 F}{\delta \Phi \delta D} = 2\{\frac{\delta[P^*\Lambda'(\Sigma^{-1} - W)\Lambda]}{\delta \Phi}(I \otimes\otimes B\Phi) + \frac{\delta B\Phi}{\delta \Phi}[\Lambda'(\Sigma^{-1} - W)\Lambda P^{*'} \otimes\otimes I]\}$$

$$= 2L_{r^2}[B'\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\otimes B'\Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ B'\Lambda' W\Lambda P^{*\prime} \otimes\!\otimes B'\Lambda'\Sigma^{-1}\Lambda B\Phi + B'\Lambda'(\Sigma^{-1} - W)\Lambda P^{*\prime} \otimes\!\otimes \Phi] \quad .$$

Using essentially the same steps in conjunction with Equation (3.23) we have

$$\frac{\delta^2 F}{\delta\Theta\delta D} = 2\,\frac{\delta(\Sigma^{-1} - W)}{\delta\Theta}\,(\Lambda P^{*\prime} \otimes\!\otimes \Lambda B\Phi) \quad ,$$

$$= 2\,\frac{\delta\Theta}{\delta\Theta}\,[\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\otimes \Lambda'(W - \Sigma^{-1})\Lambda B\Phi$$

$$+ \Lambda' W\Lambda P^{*\prime} \otimes\!\otimes \Lambda'\Sigma^{-1}\Lambda B\Phi] \quad ,$$

and, hence,

$$\text{(3.33)} \qquad \frac{\delta^2 F}{\delta\theta\delta D} = 2R[\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\otimes \Lambda'(W - \Sigma^{-1})\Lambda B\Phi + \Lambda' W\Lambda P^{*} \otimes\!\otimes \Lambda'\Sigma^{-1}\Lambda B\Phi] \quad .$$

The derivation of

$$\text{(3.34)} \qquad \frac{\delta^2 F}{\delta\Psi\delta D} = 2D(S_1)L^{*}[\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\otimes (W - \Sigma^{-1})\Lambda B\Phi + W\Lambda P^{*} \otimes\!\otimes \Sigma^{-1}\Lambda B\Phi] \quad ,$$

is by now obvious.

Rewrite Equation (3.11) as

$$\frac{\delta F}{\delta\Psi} = B'\Lambda'(\Sigma^{-1} - W)\Lambda B + B'\Lambda'(\Sigma^{-1} - W)\Lambda B - I * B'\Lambda'(\Sigma^{-1} - W)\Lambda B \quad .$$

This derivative is of the form

$$X + X' - I * X \quad .$$

Since

$$(3.35) \qquad \frac{\delta(X + X' - I * X)}{\delta X} = (I_{r^2} + E_{r^2} - J_{r^2}) \quad ,$$

$$= L_{r^2} \quad ,$$

where the matrices I, E, J and L are defined in the Appendix, it follows that

$$\frac{\delta^2 F}{\delta\Phi\delta\Phi} = \{\frac{\delta[B'\Lambda'(\Sigma^{-1} - W)\Lambda B]}{\delta\Phi}\}$$

$$\{\frac{\delta[B'\Lambda'(\Sigma^{-1} - W)\Lambda B + B'\Lambda'(\Sigma^{-1} - W)\Lambda B - I * B'\Lambda'(\Sigma^{-1} - W)\Lambda B]}{\delta B'\Lambda'(\Sigma^{-1} - W)\Lambda B}\} \quad ,$$

$$= \frac{\delta[B'\Lambda'(\Sigma^{-1} - W)\Lambda B]}{\delta\Phi} L_{r^2} \quad ,$$

$$= \frac{\delta(\Sigma^{-1} - W)}{\delta\Phi} (\Lambda B \otimes\otimes \Lambda B) L_{r^2} \quad .$$

Using Equation (3.22), we have

$$(3.36)\qquad \frac{\delta^2 F}{\delta\Phi\delta\Phi} = L_{r^2}[B'\Lambda'\Sigma^{-1}\Lambda B \otimes\otimes B'\Lambda'(W - \Sigma^{-1})\Lambda B$$

$$+ B'\Lambda' W\Lambda B \otimes\otimes B'\Lambda'\Sigma^{-1}\Lambda B]L_{r^2} \quad .$$

Further, with the aid of Equation (3.23)

$$\frac{\delta^2 F}{\delta\Theta\delta\Phi} = \frac{\delta(\Sigma^{-1} - W)}{\delta\Theta} (\Lambda B \otimes\otimes \Lambda B)L_{r^2} \quad ,$$

$$= \frac{\delta\Theta}{\delta\Theta} [\Lambda'\Sigma^{-1}\Lambda B \otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda B$$

$$+ \Lambda' W\Lambda B \otimes\otimes \Lambda'\Sigma^{-1}\Lambda B]L_{r^2} \quad .$$

Since

$$\frac{\delta^2 F}{\delta\theta\delta\Phi} = \frac{\delta\Theta}{\delta\theta} \frac{\delta^2 F}{\delta\Theta\delta\Phi} \quad ,$$

using Equation (3.14)

$$(3.37)\qquad \frac{\delta^2 F}{\delta\theta\delta\Phi} = R[\Lambda'\Sigma^{-1}\Lambda B \otimes\otimes \Lambda'(W - \Sigma^{-1})\Lambda B + \Lambda' W\Lambda B \otimes\otimes \Lambda'\Sigma^{-1}\Lambda B]L_{r^2} \quad .$$

Using similar steps

$$\frac{\delta^2 F}{\delta\Psi\delta\Phi} = \frac{\delta(\Sigma^{-1} - W)}{\delta\Psi} (\Lambda B \otimes\otimes \Lambda B)L_{r^2} \quad ,$$

which upon substituting (3.26) becomes

(3.38) $$\frac{\delta^2 F}{\delta\Psi\delta\Phi} = D(S_1)L*[\Sigma^{-1}\Lambda B \otimes\otimes (W - \Sigma^{-1})\Lambda B + W\Lambda B \otimes\otimes \Sigma^{-1}\Lambda B]L_{r^2} \quad .$$

Rewriting Equation (3.15) as

$$\frac{\delta F}{\delta\theta} = \mathrm{Vec}[\Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 + \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 - I \quad \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2] \quad ,$$

then using the results of Equation (3.35) we obtain

(3.39) $$\frac{\delta^2 F}{\delta\theta\delta\theta} = \frac{\delta[\Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2]}{\delta\theta} L_{r^2} \quad ,$$

$$= \frac{\delta(\Sigma^{22} - W_{22})}{\delta\theta} (\Lambda_2 \otimes \Lambda_2)L_{r^2} \quad .$$

An expression for $\delta(\Sigma^{22} - W_{22})/\delta\theta$ may be obtained by locating the submatric element, $\delta(\Sigma^{22} - W_{22})/\delta\theta$, in

$$\frac{\delta(\Sigma^{-1} - W)}{\delta\Theta} = \frac{\delta\Theta}{\delta\Theta} (\Lambda'\Sigma^{-1} \otimes\otimes \Lambda'(W - \Sigma^{-1}) + \Lambda'W \otimes\otimes \Lambda'\Sigma^{-1}) \quad ,$$

and evaluating this element. Now $\delta(\Sigma^{-1} - W)/\delta\Theta$ is a $(4r^2 x 4p^2)$ matrix with $\delta(\Sigma^{22} - W_{22})/\delta\theta$ as the lower righthand $(r^2 x p^2)$ submatric element. Therefore, it is necessary to obtain an expression for the lower righthand submatric element. From the definition of the Double Kronecker Product, we obtain

(3.40) $$\Lambda_2'\Sigma^{22} \otimes \Lambda_2'(W_{22} - \Sigma^{22}) + \Lambda_2'W_{22} \otimes \Lambda_2'\Sigma^{22} \quad ,$$

as the desired expression. Therefore,

$$(3.41)\qquad \frac{\delta(\Sigma^{22} - W_{22})}{\delta\theta} = \frac{\delta\theta}{\delta\theta}\,[\Lambda_2'\Sigma^{22} \otimes \Lambda_2'(W_{22} - \Sigma^{22}) + \Lambda_2'W_{22} \otimes \Lambda_2'\Sigma_{22}] \quad .$$

Substituting (3.41) in (3.39) we obtain

$$(3.42)\qquad \frac{\delta^2 F}{\delta\theta\delta\theta} = L_{r^2}[\Lambda_2'\Sigma^{22} \otimes \Lambda_2'(W_{22} - \Sigma^{22}) + \Lambda_2'W_{22} \otimes \Lambda_2'\Sigma_{22}]L_{r^2} \quad .$$

In a similar fashion, with the aid of Equation (3.24), we obtain

$$(3.43)\qquad \frac{\delta^2 F}{\delta\Psi\delta\theta} = D(S_1)L_{p^2}[\Sigma^{22}\Lambda_2 \otimes (W_{22} - \Sigma^{22})\Lambda_2 + W_{22}\Lambda_2 \otimes \Sigma^{22}\Lambda_2]L_{r^2} \quad .$$

Rewrite (3.16) as

$$\frac{\delta F}{\delta\Psi} = S_1 * [(\Sigma^{-1} - W) + (\Sigma^{-1} - W) - I * (\Sigma^{-1} - W)] \quad .$$

Then,

$$\frac{\delta^2 F}{\delta\Psi\delta\Psi} = \frac{\delta(\Sigma^{-1} - W)}{\delta\Psi}\,\frac{\delta\{S_1 * [(\Sigma^{-1} - W) + (\Sigma^{-1} - W)' - I * (\Sigma^{-1} - W)]\}}{\delta(\Sigma^{-1} - W)} \quad .$$

Now

$$(\Sigma^{-1} - W) + (\Sigma^{-1} - W)' - I * (\Sigma^{-1} - W) \quad ,$$

is of the form

$$X + X' - I * X \quad ,$$

where X (2px2p) is a partitioned matrix with X_{ii} symmetric and $X_{ij} = X'_{ji}$. It has been shown (McDonald and Swaminathan, 1972) that

$$\text{(3.44)} \qquad \frac{\delta(X + X' - I \quad X)}{\delta X} = I^* + E^* - J^* \quad ,$$

$$= L^* \quad ,$$

where

$$I^* = \begin{bmatrix} I_{p^2} & 0 & 0 & 0 \\ 0 & I_{p^2} & 0 & 0 \\ 0 & 0 & I_{p^2} & 0 \\ 0 & 0 & 0 & I_{p^2} \end{bmatrix} ,$$

$$E^* = \begin{bmatrix} E_{p^2} & 0 & 0 & 0 \\ 0 & 0 & E_{p^2} & 0 \\ 0 & E_{p^2} & 0 & 0 \\ 0 & 0 & 0 & E_{p^2} \end{bmatrix} ,$$

$$J^* = \begin{bmatrix} J_{p^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & J_{p^2} \end{bmatrix} ,$$

and L* has been defined previously. The matrices I*, E*, J* and L* are the partitioned counterparts of I, E, J and L respectively. Using Equations (3.24) and (3.44)

$$\text{(3.45)} \qquad \frac{\delta^2 F}{\delta\Psi\delta\Psi} = D(S_1)L^*[\Sigma^{-1} \otimes\otimes (W - \Sigma^{-1}) + W \otimes\otimes \Sigma^{-1}]L^*D(S_1) \quad .$$

Taking account of the form of $\delta F/\delta\Psi_{ij\ \text{diag}}$, Equation (3.45) may be expressed in terms of its submatric elements as follows:

$$\text{(3.46a)} \qquad \frac{\delta^2 F}{\delta\Psi_{11\ \text{diag}}\delta\Psi_{11\ \text{diag}}} = \Sigma^{11} * (W_{11} - \Sigma^{11}) + (W_{11} * \Sigma^{11}) \quad ,$$

$$\text{(3.46b)} \qquad \frac{\delta^2 F}{\delta\Psi_{12\ \text{diag}}\delta\Psi_{11\ \text{diag}}} = \Sigma^{11} * (W_{12} - \Sigma^{12}) + (W_{11} * \Sigma^{12}) \quad ,$$

$$\text{(3.46c)} \qquad \frac{\delta^2 F}{\delta\Psi_{22\ \text{diag}}\delta\Psi_{11\ \text{diag}}} = \Sigma^{12} * (W_{12} - \Sigma^{12}) + (W_{12} * \Sigma^{12}) \quad ,$$

$$\text{(3.46d)} \qquad \frac{\delta^2 F}{\delta\Psi_{11\ \text{diag}}\delta\Psi_{12\ \text{diag}}} = 4[(\Sigma^{11} * (W_{21} - \Sigma^{21}) + (W_{11} * \Sigma^{21})] \quad ,$$

(3.46e) $$\frac{\delta^2 F}{\delta\Psi_{12\ diag}\,\delta\Psi_{12\ diag}} = 2[(\Sigma^{11} * (W_{22} - \Sigma^{22}) + (W_{11} * \Sigma^{22})]$$

$$+ 2[\Sigma^{21} * (W_{12} - \Sigma^{12}) + (W_{12} * \Sigma^{12})] \quad ,$$

(3.46f) $$\frac{\delta^2 F}{\delta\Psi_{22\ diag}\,\delta\Psi_{12\ diag}} = 4[\Sigma^{12} * (W_{22} - \Sigma^{22}) + (W_{12} * \Sigma^{22})] \quad ,$$

(3.46g) $$\frac{\delta^2 F}{\delta\Psi_{11\ diag}\,\delta\Psi_{22\ diag}} = \Sigma^{21} * (W_{21} - \Sigma^{21}) + (W_{21} * \Sigma^{21}) \quad ,$$

(3.46h) $$\frac{\delta^2 F}{\delta\Psi_{12\ diag}\,\delta\Psi_{22\ diag}} = \Sigma^{21} * (W_{22} - \Sigma^{22}) + W_{21} * \Sigma^{22}) \quad ,$$

and

(3.46i) $$\frac{\delta^2 F}{\delta\Psi_{22\ diag}\,\delta\Psi_{22\ diag}} = \Sigma^{22} * (W_{22} - \Sigma^{22}) + (W_{22} * \Sigma^{22}) \quad .$$

3.3 Effect of Change in the Unit of Measurement

In this section the effect of a change in the units of measurement of the elements of the vector $\underline{y}$ is considered. Examining the effect of such a change is of importance, since the unit of measurement for educational and psychological measurements is generally considered to be arbitrary. The issues that will be explored are the effect of the scale transformation on the parameters and the effect of the scale transformation on the estimates of the parameters.

Changing the units of measurement means multiplying each component of $\underline{y}$ by a constant. This change may be represented by the equation

$$\underline{z} = V\underline{y} \quad ,$$

where V is a (2px2p) diagonal matrix of constants. The diagonal submatrices, V_1 (pxp) and V_2 (pxp), contain the only non-zero elements in V. The effect of the scale transformation may be clarified by observing that

$$(3.47) \qquad F(\underline{z}\underline{z}') = \Sigma^{(z)} = V\Sigma V \quad ,$$

$$= V\Lambda(B\Phi B' + \Theta)\Lambda'V' + V\Psi V' \quad .$$

Clearly B, Φ and Θ have not been effected by the scale transformation and we can write (3.47) as

$$(3.48) \qquad \Sigma^{(z)} = \Lambda^{(z)}(B\Phi B' + \Theta)\Lambda^{(z)} + \Psi^{(z)} \quad .$$

Assume that ψ_{11} and ψ_{22} are identified. From Equations (3.47) and (3.48 and the identification of ψ_{11} and ψ_{22} we find that

$$(3.49a) \qquad \psi_{11}^{(z)} = V_1\psi_{11}V_1' \quad ,$$

(3.49b) $$\psi_{12}^{(z)} = V_1\psi_{12}V_2' \quad ,$$

(3.49c) $$\psi_{22}^{(z)} = V_2\psi_{22}V_2' \quad ,$$

(3.49d) $$\Lambda_1^{(z)}\Phi\Lambda_1^{(z)\prime} = V_1\Lambda_1\Phi\Lambda_1'V_1' \quad ,$$

(3.49e) $$\Lambda_2^{(z)}(D\Phi D' + \theta)\Lambda_2^{(z)\prime} = V_2\Lambda_2(D\Phi D' + \theta)\Lambda_2'V_2' \quad ,$$

and

(3.49f) $$\Lambda_1^{(z)}\Phi D'\Lambda_2^{(z)\prime} = V_1\Lambda_1\Phi D'\Lambda_2'V_2' \quad ,$$

but whether

(3.50a) $$\Lambda_1^{(z)} = V_1\Lambda_1 \quad ,$$

and

(3.50b) $$\Lambda_2^{(z)} = V_2\Lambda_2 \quad ,$$

depends on the identifying conditions that Λ must satisfy (Anderson and Rubin, 1956).

If Λ_1 and Λ_2 are identified, up to a post-multiplication by a diagonal matrix, by the occurrence of $r - 1$ zeros in specified places in the columns of Λ_1 and Λ_2, then $\Lambda_1^{(z)}(= V_1\Lambda_1)$ and $\Lambda_2^{(z)}(= V_2\Lambda_2)$

will have corresponding zero elements. Thus, Equations (3.50) will hold. However, if normalization rules that set one element in each column of Λ_1 and Λ_2 to unity, are introduced, then, in general, the normalization rules will not be satisfied for $\Lambda_1^{(z)}$ (= $V_1\Lambda_1$) or $\Lambda_2^{(z)}$ (= $V_2\Lambda_2$). Since normalization rules are generally arbitrary, the failure for the rule to hold is not very serious. The matrices $\Lambda_1^{(z)}$ (= $V_1\Lambda_1$) and $\Lambda_2^{(z)}$ (= $V_2\Lambda_2$) will meet the conditions for identification with a new scale chosen for the parameters.

Suppose that at least one of the r - 1 restrictions on, say, $\underset{\sim}{\lambda}_{11}$ specifies that a linear combination of the elements of $\underset{\sim}{\lambda}_{11}$ equals zero. As was noted above an example of this type of restriction is one that restricts two elements of $\underset{\sim}{\lambda}_{11}$ to be equal. Let all the restrictions on $\underset{\sim}{\lambda}_{11}$ be represented as

$$R\underset{\sim}{\lambda}_{11} = \underset{\sim}{0} \quad ,$$

the vector

$$\underset{\sim}{\lambda}_{11}^{(z)} = V\underset{\sim}{\lambda}_{11} \quad ,$$

will not, in general, satisfy the restriction

$$R\underset{\sim}{\lambda}_{11}^{(z)} = \underset{\sim}{0} \quad ,$$

and, therefore, Equations (3.50) do not obtain.

If the structure for Σ is identified under the conditions of Theorem 2.3 or Theorem 2.4, then these conditions will be satisfied for the structure of $\Sigma^{(z)}$ since V is a non-singular matrix. Therefore, Equations (3.50) will hold.

The next issue that will be examined is that of the effect of a change in the units of measurement on the maximum likelihood estimates of the parameters. Ignoring the forms of Λ, B and θ the likelihood equations may be written as

$$\text{(3.51a)} \qquad \frac{\delta F}{\delta \Lambda} = (\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda'(B\Phi B' + \Theta) = 0 \quad ,$$

$$\text{(3.51b)} \qquad \frac{\delta F}{\delta B} = \Lambda(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda' B\Phi = 0 \quad ,$$

$$\text{(3.51c)} \qquad \frac{\delta F}{\delta \Phi} = 2B'\Lambda'(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda B - I * B'\Lambda'(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda B = 0 \quad ,$$

$$\text{(3.51d)} \qquad \frac{\delta F}{\delta \Theta} = 2\Lambda'(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda - I * \Lambda'(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})\Lambda = 0 \quad ,$$

and

$$\text{(3.51e)} \qquad \frac{\delta F}{\delta \Psi} = S_1\, [2(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1}) - I * (\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})] = 0 \quad .$$

That is, Equations (3.51) are the likelihood equations if we ignore the fact that certain elements of the matrices Λ, B and Θ, given by Equations (2.9c), (2.9d) and (2.9f), are zero, and certain elements of B are unity. It is easily seen that using Equations (3.51) rather than Equations (3.18) in the following proof, simplifies the expressions but does not invalidate the results.

Since

$$\Sigma^{(z)} = V\Sigma V \quad ,$$

and

$$S^{(z)} = VSV \quad ,$$

Equations (3.51) may be written, based on the vector $\underline{z}$ as

$$(3.52a) \qquad (V^{-1}\Sigma^{-1}V^{-1} - V^{-1}\Sigma^{-1}V^{-1}VSVV^{-1}\Sigma^{-1}V^{-1})\Lambda^{(z)}(B\Phi B' + \Theta) = 0 \quad ,$$

$$(3.52b) \qquad \Lambda^{(z)\prime}(V^{-1}\Sigma^{-1}V^{-1} - V^{-1}\Sigma^{-1}V^{-1}VSVV^{-1}\Sigma^{-1}V^{-1})\Lambda^{(z)}B\Phi = 0 \quad ,$$

$$(3.52c) \qquad B'\Lambda^{(z)\prime}(V^{-1}\Sigma^{-1}V^{-1} - V^{-1}\Sigma^{-1}V^{-1}VSVV^{-1}\Sigma^{-1}V^{-1})\Lambda^{(z)}B = 0 \quad ,$$

$$(3.52d) \qquad \Lambda^{(z)\prime}(V^{-1}\Sigma^{-1}V^{-1} - V^{-1}\Sigma^{-1}V^{-1}VSVV^{-1}\Sigma^{-1}V^{-1})\Lambda^{(z)} = 0 \quad ,$$

and

$$(3.52e) \qquad S_1 * (V^{-1}\Sigma^{-1}V^{-1} - V^{-1}\Sigma^{-1}V^{-1}VSVV^{-1}\Sigma^{-1}V^{-1}) = 0 \quad .$$

Pre-multiply Equation (3.52a) by V and pre and post-multiply

Equation (3.52e) by V to obtain

$$(3.53a) \qquad (\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})V^{-1}\Lambda^{(z)}(B\Phi B' + \Theta) = 0 \quad ,$$

$$(3.53b) \qquad \Lambda^{(z)\prime}V^{-1}(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})V^{-1}\Lambda^{(z)}B\Phi = 0 \quad ,$$

$$(3.53c) \qquad B'\Lambda^{(z)\prime}V^{-1}(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})V^{-1}\Lambda^{(z)}B = 0 \quad ,$$

$$(3.53d) \qquad \Lambda^{(z)\prime}V^{-1}(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1})V^{-1}\Lambda^{(z)} = 0 \quad ,$$

and

$$(3.53e) \qquad S_1 * (\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1}) = 0 \quad .$$

Comparing Equations (3.53) with (3.51) it is clear that we may take

$$\Lambda^{(z)} = V\Lambda \quad ,$$

conditional upon the requirement that both $\Lambda^{(z)}$ (= $V\Lambda$) and Λ satisfy the same conditions for identification. Thus, maximum likelihood estimators of the parameters are effected by the scale transformation in the same way the parameters are.

Suppose that the matrix V contains as its diagonal elements reciprocals of the population variances of the random vector $\underline{y}$. The matrix $\Sigma^{(z)}$ is then the population correlation matrix, and the parameter matrices $\Lambda_1^{(z)}$, $\Lambda_2^{(z)}$, $\psi_{11}^{(z)}$, $\psi_{12}^{(z)}$, and $\psi_{22}^{(z)}$ are parts of the solution to the model in terms of the correlation matrix. It is of interest to inquire what the maximum likelihood estimates of these matrices are.

The problem that arises in estimating these parameter matrices is that the sample estimate of the correlation matrix does not follow the Wishart distribution and so Equations (3.18) may not be used to estimate the parameters. Moreover, in this case, the immediately preceeding results concerning the effect of a scale transformation on the estimates cannot be utilized, since generally the population values for the variances of the variables are unknown. However, the following theorem, which permits the values of these parameters to be estimated from the parameters estimated by Equations (3.18) on the basis of the covariance matrix, is well known.

<u>Theorem 3.1</u>: If on the basis of a given sample the vector $\hat{\underline{\Theta}}$ is the maximum likelihood estimate of the parameter vector $\underline{\Theta}$, then $f_1(\hat{\underline{\Theta}})$, $f_2(\hat{\underline{\Theta}})$, ... , $f_m(\hat{\underline{\Theta}})$ are maximum likelihood estimates of $f_1(\underline{\Theta})$, ... , $f_m(\underline{\Theta})$. If the estimates of $\underline{\Theta}$ are unique then the estimates of $f_1(\underline{\Theta})$, ... , $f_m(\underline{\Theta})$ are unique (Anderson, 1958).

The transformations given by (3.49a), (3.49b), (3.49c), (3.50a) and (3.50b) are one to one. Therefore, the maximum likelihood estimates of $\Lambda_1^{(z)}$, $\Lambda_2^{(z)}$, $\psi_{11}^{(z)}$, $\psi_{12}^{(z)}$ and $\psi_{22}^{(z)}$, which in this case are the estimates of the structural parameters for the correlation matrix, can be obtained by replacing the population matrices by their sample analogues in Equations (3.49a), (3.49b), (3.49c), (3.50a) and (3.50b).

That is, the diagonal elements of S are substituted for V and the sample estimates $\hat{\Lambda}_1$, $\hat{\Lambda}_2$, $\hat{\psi}_{11}$, $\hat{\psi}_{12}$, and $\hat{\psi}_{22}$, based on the covariance matrix, are substituted for Λ_1, Λ_2, ψ_{11}, ψ_{12}, and ψ_{22}. Again, however, the conclusion is conditional upon the requirement that $\Lambda^{(z)}$ (= $V\Lambda$) and Λ satisfy the same conditions for identification.

CHAPTER IV

TESTS OF HYPOTHESES FOR THE LONGITUDINAL FACTOR ANALYSIS MODEL

4.0 Introduction

In this chapter methods for testing hypotheses about the parameters of the structural model are presented. In section 4.1, the likelihood ratio test of the adequacy of the structural model is presented. Tests of the various hypotheses about parametric invariance are derived in section 4.2. Following Joreskog (1970) and Swaminathan (1972, 1973), these hypotheses are expressed as hypotheses about the parametric structure of Σ, and tests of these hypotheses are derived using the likelihood ratio criterion. In section 4.3, methods for testing hypotheses about a given element or group of elements of a parameter matrix are given. These methods, based on the well known asymptotic properties of maximum likelihood estimates, require the evaluation of the expected value of the Hessian, at the minimum of the function given by Equation (3.3). Hence, expressions for the expected values of the second derivatives are given in this section.

4.1 The Likelihood Ratio Test of the Adequacy of the Structural Model

The hypothesis about the adequacy of the structural model may be expressed as a hypothesis about the parametric structure of Σ, and may be tested using the likelihood ratio criterion. In this section, the likelihood ratio test is first discussed in general. The application of the likelihood ratio test to the problem of testing the adequacy of the structural model is then presented.

Let H_o be a hypothesis concerning the parametric structure of Σ. That is, a hypothesis that the elements of Σ belong to the set of points in the parameter space for which the hypothesized parametric structure exists. Denote this set of points as ω. Let H_A be any alternative hypothesis that Σ is contained in a region of the parameter space, Ω, such that $\omega \subseteq \Omega$. It is necessary to construct a test of the hypothesis

$$H_o : \Sigma \in \omega \quad ,$$

against the alternative

$$H_A : \Sigma \in \Omega \quad .$$

Let L_ω be the maximum of the logarithm of the likelihood function,

$$\log L = -\frac{n}{2}\,[\log|\Sigma| + tr(S\Sigma^{-1})] \quad ,$$

under H_o and L_Ω be the maximum of logL under H_A. Then obviously

$$L_\omega \leq L_\Omega \quad ,$$

since the maximum of a function in a restricted space must be less than or equal to the maximum of the function in an unrestricted space. Thus, the likelihood ratio criterion

$$\log\lambda = L_\omega - L_\Omega \quad ,$$

is negative, i.e., $0<\lambda\leq 1$. It is well known that asymptotically

$$-2\log\lambda$$

is distributed as χ^2 with degrees of freedom equal to the difference in the number of parameters estimated under H_A and H_o. If the probability that a χ^2 random variable exceeds the likelihood ratio criterion, $-2\log\lambda$, is small, then H_A is favored, and we conclude that Σ has the less restrictive form.

Intuitively speaking, the likelihood ratio criterion provides a test of the similarity of the estimates of Σ obtained under H_o and H_A. The estimate of Σ, $\hat{\Sigma}_\omega$, obtained under H_o must conform to the parametric structure hypothesized under H_o. The estimate of Σ, $\hat{\Sigma}_\Omega$, under H_A conforms to a structure that is less restrictive than the structure hypothesized under H_o. However,

$$\hat{\Sigma}_\omega = \hat{\Sigma}_\Omega \quad ,$$

except for differences attributable to random variation, implies that $\hat{\Sigma}_\Omega$ does conform to the more restrictive structure given by H_o. Thus, it is reasonable to accept H_o. That is, to consider Σ to have the parametric structure given by H_o.

In order to test the adequacy of the structural model, we follow the procedure outlined above. The test of the adequacy of the model is the test of the hypothesis that Σ has the form given by Equation (2.10). That is, the hypothesis

$$(4.1) \qquad H_o^{(1)} : \Sigma = \Lambda(B\Phi B' + \Theta)\Lambda' + S_1 * \Psi \quad ,$$

which is tested against

$$(4.2) \qquad H_A^{(1)} : \Sigma \text{ is any positive definite matrix.}$$

It is well known that the maximum L_Ω under H_A is obtained when

$$\hat{\Sigma} = S \quad ,$$

where S is the sample dispersion matrix. Thus,

$$\log L = -\frac{n}{2}\,[\log|\hat{\Sigma}| + \mathrm{Tr}(S\hat{\Sigma}^{-1})] \quad ,$$

evaluated when $\hat{\Sigma} = S$, yields

$$L_\Omega = -\frac{n}{2}\,(\log|S| + 2p) \quad .$$

Further,

$$L_\omega = -\frac{n}{2}\,[\log|\hat{\Sigma}| + \mathrm{Tr}(S\hat{\Sigma}^{-1})] \quad ,$$

where

$$\hat{\Sigma} = \hat{\Lambda}(\hat{B}\hat{\Phi}\hat{B}' + \hat{\Theta})\hat{\Lambda}' + S_1 * \hat{\Psi} \quad .$$

The matrices $\hat{\Lambda}$, $\hat{B}$, $\hat{\Phi}$, $\hat{\Theta}$ and $\hat{\Psi}$ are obtained by solving Equations (3.18).

Therefore,

$$(4.3) \qquad L_\omega - L_\Omega = -\frac{n}{2}\,[\log|\hat{\Sigma}| + \mathrm{Tr}(S\hat{\Sigma}^{-1}) - \log|S| - 2p] \quad .$$

Whence,

$$-2\log\lambda = n[\log|\hat{\Sigma}| + \mathrm{Tr}(S\hat{\Sigma}^{-1}) - \log|S| - 2p] \quad ,$$

which simplifies to

$$-2\log\lambda = n[\text{minimum of } F(\Lambda, B, \Phi, \Theta, \Psi)] \quad .$$

Thus, the likelihood ratio criterion is equal to n times the minimum of F when the parameters are free to vary (except for the restrictions required for estimation). Denote this likelihood ratio by nF_{min}. If n is large, the criterion nF_{min} has the χ^2 distribution with degrees of freedom ν, where ν is given by

$$\nu = p(2p + 1) - (2pr + 3p + r) \quad ,$$

$$= 2p^2 - (2pr + 3p + r) \quad .$$

As was mentioned in section 2.2, the quantity ν is the quantity involved in the inequality

$$\nu = p(2p + 1) - (2pr + 3p + r) > 0 \quad ,$$

or

$$r \leq 2p(p - 1)/2p + 1 \quad ,$$

which was suggested by the comparison approach as a minimal condition for identification. From a practical point of view, if the inequality does not obtain, then the hypothesis given by Equation (4.1) cannot be tested. In this case, there would be negative degrees of freedom associated with the likelihood ratio test. It is interesting to note that the condition

$$\nu \geq 0 \quad ,$$

which is considered to be a minimal condition for identification, is also a necessary condition for testing the hypothesis of the existence of the structure.

4.2 Likelihood Ratio Tests of Parametric Invariance

In this section likelihood ratio tests of the invariance of various parameters of the model are derived. However, before presenting these tests, another hypothesis that appears to have implications for the study of parametric invariance, is discussed. This is the hypothesis of equality of the first and second occasion covariance matrices, that is, the hypothesis

$$(4.4) \qquad H_o^{(2)} : \Sigma_{11} = \Sigma_{22} = \Sigma_0 \quad ,$$

which is tested against the alternative

$$(4.5) \qquad H_A^{(2)} : \Sigma \text{ is any positive definite matrix.}$$

In terms of the structural model, the first occasion covariance matrix is

$$\Sigma_{11} = \Lambda_1 \Phi \Lambda_1' + \psi_{11} \quad ,$$

while the second occasion covariance matrix is

$$\Sigma_{22} = \Lambda_2 (D\Phi D' + \theta)\Lambda_2' + \psi_{22} \quad .$$

Suppose that the covariance matrices for the two occasions are equal. Then

$$\Sigma_{11} = \Sigma_{22} \quad ,$$

$$\Lambda_1 \Phi \Lambda_1' + \psi_{11} = \Lambda_2 (D \Phi D' + \theta) \Lambda_2' + \psi_{22} \quad .$$

If the structure is identified, then Λ_1, Λ_2, D, Φ, θ, ψ_{11} and ψ_{22} are unique, and

$$\Lambda_1 = \Lambda_2 \quad ,$$

$$\Phi = D \Phi D' + \theta \quad ,$$

and

$$\psi_{11} = \psi_{22} \quad .$$

Hence, $\Sigma_{11} = \Sigma_{22}$ implies that the factor patterns are invariant, the first and second occasion factor score covariance matrices are equal and the first and second occasion uniqueness matrices are equal. The implication, however, is conditional upon the identification of the structure.

It is important to realize that accepting the hypothesis of equality of the covariance matrices for the two occasions does not imply that the factor scores are invariant over occasions. Rather, it implies that the joint distribition of the factors remains the same over occasions. Also, rejection of the hypothesis of equality of the covariance matrices does not necessarily imply that all the parameters vary over occasions. Therefore, regardless of the acceptance or rejection of the hypothesis, tests of additional hypotheses are required to examine the invariance of all the parameters.

The test of equality of the observed score covariance matrices has been given by McDonald and Swaminathan (1972). The hypothesis

$$H_o^{(2)} : \Sigma_{11} = \Sigma_{22} = \Sigma_0 \quad ,$$

implies that Σ has the form

$$\Sigma = \begin{bmatrix} \Sigma_0 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_0 \end{bmatrix} \quad .$$

The test statistic can be shown to be,

$$-2\log\lambda = n[\log|\hat{\Sigma}| + \mathrm{Tr}(S\hat{\Sigma}^{-1}) - \log|S| - 2p] \quad ,$$

where $\hat{\Sigma}$ is the solution to the system of equations

$$\frac{\delta\Sigma}{\delta\Sigma}(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1}) = 0 \quad ,$$

where

$$\frac{\delta\Sigma}{\delta\Sigma} = \begin{bmatrix} L_{p^2} & 0 & 0 & L_{p^2} \\ 0 & I_{p^2} & E_{p^2} & 0 \\ 0 & E_{p^2} & I_{p^2} & 0 \\ L_{p^2} & 0 & 0 & L_{p^2} \end{bmatrix} .$$

After simplification the system of equations becomes

$$2(\Sigma^{11} - W_{11} + \Sigma^{22} - W_{22}) - I * (\Sigma^{11} - W_{11} + \Sigma^{22} - W_{22}) = 0 \quad ,$$

$$2(\Sigma^{12} - W_{12}) = 0 \quad .$$

Returning to the problem of testing hypotheses about the invariance of the parameters of the model, assume that the adequacy of the model has been tested and the model is considered to fit the covariance matrix satisfactorily. The investigator may now proceed to test the various hypotheses concerned with parametric invariance.

One such hypothesis is the hypothesis that the factor patterns are invariant over occasions. This is the hypothesis

$$H_o^{(3)} : \Sigma = \Lambda(B\Phi B' + \Theta)\Lambda' + S_1 * \Psi \mid \Lambda_1 = \Lambda_2 = \Lambda_0 \quad ,$$

which is tested against the alternative

$$H_A^{(3)} : \Sigma = \Lambda(B\Phi B' + \Theta)\Lambda' + S_1 * \Psi \quad .$$

Hence, we are testing the hypothesis that Σ has the form given by the structural model, with the additional condition that $\Lambda_1 = \Lambda_2$. This hypothesis is tested against the alternative that Σ has the form given by the structural model. It will be more convenient to write the null and alternate hypotheses as

$$(4.6) \qquad H_o^{(3a)} : \Lambda_1 = \Lambda_2 = \Lambda_0 \quad ,$$

versus

$$(4.7) \qquad H_A^{(3a)} : \Lambda_1 \neq \Lambda_2 \quad .$$

In order to test the hypothesis the value of Equation (3.3), with $\hat{\Sigma}$ estimated under the hypothesis that $\Lambda_1 = \Lambda_2$, must be computed. This may be accomplished by substituting

$$\frac{\delta F}{\delta \Lambda} = 2[I \ I] \ [(\Sigma^{-1} - W)\Lambda\Gamma]_{diag} \quad ,$$

$$= 2[(\Sigma^{-1} - W)\Lambda\Gamma]_{11} + 2[(\Sigma^{-1} - W)\Lambda\Gamma]_{22} \quad ,$$

in the lefthand side of Equation (3.18a), solving Equations (3.18) for $\hat{\Lambda}_0$, $\hat{D}$, $\hat{\Phi}$, $\hat{\theta}$ and $\hat{\Psi}$, and computing the value of Equation (3.3).

Denote the resulting value of (3.3) by $F_{min}(\Lambda_0)$. Then the likelihood ratio criterion is given by

$$-2\log\lambda = n[F_{min}(\Lambda_0) - F_{min}] \quad ,$$

and is approximately distributed as χ^2 with $r(p - r)$ degrees of freedom.

Regardless of whether the hypothesis given by (4.6) is accepted or rejected, the investigator may be interested in testing the hypothesis that the factor structure is invariant. This is the hypothesis that the covariance matrix between the variables and factors is invariant over occasions. If invariance holds, then

$$E(\underline{y}_1\underline{x}_1') = E(\underline{y}_2\underline{x}_2') \quad ,$$

and, therefore,

$$\Lambda_1\Phi = \Lambda_2(D\Phi D' + \theta) \quad , \tag{4.8}$$

$$\Lambda_1 = \Lambda_2(D\Phi D' + \theta)\Phi^{-1} \quad .$$

Another hypothesis that may be of interest concerns whether

$$\Delta_p^{-1}\Lambda_1\Phi\Delta_r^{-1} = \Lambda_2(D\Phi D' + \theta) \quad ,$$

or equivalently

$$\Lambda_1 = \Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r\Phi^{-1} \quad , \tag{4.9}$$

where Δ_p (pxp) and Δ_r (rxr) are diagonal matrices. The hypothesis is concerned with the question of whether the correlation matrix between factors and observed scores is invariant.

In order to test these hypotheses, it is necessary to estimate the model under the side conditions on Λ_1 given by Equations (4.8) and (4.9). For estimation under the condition given by Equation (4.9), the derivatives $\delta F/\delta\Lambda_2$, $\delta F/\delta\Delta_p$, $\delta F/\delta\Delta_r$, $\delta F/\delta\Phi$, $\delta F/\delta B$, $\delta F/\delta\theta$ and $\delta F/\delta\Psi$ are required. Since Equation (4.8) is essentially a specialization of Equation (4.9) with both Δ_p and Δ_r as identity matrices, only these derivatives will be required.

In order to obtain the required derivatives, we first let

$$A = \begin{bmatrix} \Delta_p & 0 \\ 0 & I \end{bmatrix} ,$$

$$C = \begin{bmatrix} \Lambda_2 & 0 \\ 0 & \Lambda_2 \end{bmatrix} ,$$

$$G = \begin{bmatrix} D\Phi D' + \theta & 0 \\ 0 & I \end{bmatrix} ,$$

$$K = \begin{bmatrix} \Delta_r & 0 \\ 0 & I \end{bmatrix} ,$$

$$M = \begin{bmatrix} \Phi^{-1} & 0 \\ 0 & I \end{bmatrix} ,$$

and rewrite the model as

$$(4.10) \qquad \Sigma = ACGKM(B\Phi B' + \Theta)MKGC'A + S_1 \quad \Psi \quad .$$

Using steps similar to those utilized in obtaining the derivatives in the preceeding chapter, we obtain the following derivatives:

$$(4.11) \qquad \frac{\delta F}{\delta \Lambda_2} = 2\mathrm{Vec}[\Delta_p(\Sigma^{11} - W_{11})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r\Phi^{-1}\Delta_r(D\Phi D' + \theta)]$$

$$+ 2\ \mathrm{Vec}[\Delta_p(\Sigma^{12} - W_{12})\Lambda_2 D\Delta_r(D\Phi D' + \theta)]$$

$$+ 2\ \mathrm{Vec}[(\Sigma^{21} - W_{21})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r D']$$

$$+ 2\ \mathrm{Vec}[(\Sigma^{22} - W_{22})\Lambda_2(D\Phi D' + \theta)] \quad ,$$

$$(4.12) \qquad \frac{\delta F}{\delta \Delta_p} = 2\mathrm{Vec}[(\Sigma^{11} - W_{11})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r\Phi^{-1}\Delta_r(D\Phi D' + \theta)\Lambda_2'$$

$$+ (\Sigma^{12} - W_{12})\Lambda_2 D\Delta_r(D\Phi D' + \theta)\Lambda_2]_{diag} \quad ,$$

$$(4.13) \qquad \frac{\delta F}{\delta \Delta_r} = 2\mathrm{Vec}[(D\Phi D' + \theta)\Lambda_2'\Delta_p(\Sigma^{11} - W_{11})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r\Phi^{-1}$$

$$+ (D\Phi D' + \theta)\Lambda_2'\Delta_p(\Sigma^{12} - W_{12})\Lambda_2 D]_{diag} \quad ,$$

(4.14) $$\frac{\delta F}{\delta D} = 2\mathrm{Vec}[\Lambda_2'\Delta_p(\Sigma^{11} - W_{11})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r\Phi^{-1}\Delta_r D\Phi$$

$$+ \Delta_r\Phi^{-1}\Delta_r(D\Phi D' + \theta)\Lambda_2'\Delta_p(\Sigma^{11} - W_{11})\Delta_p\Lambda_2 D\Phi$$

$$+ \Lambda_2'\Delta_p(\Sigma^{12} - W_{12})\Lambda_2 D\Delta_r D\Phi + \Delta_r\theta\Lambda_2'\Delta_p(\Sigma^{12} - W_{12})\Lambda_2$$

$$+ \Delta_r D'\Lambda_2'(\Sigma^{21} - W_{21})\Delta_r\Lambda_2 D\Phi + \Lambda_2'(\Sigma^{21} - W_{21})\Delta_p\Lambda_2' D\Phi D'\Delta_r$$

$$+ \Lambda_2(\Sigma^{22} - W_{22})\Lambda_2 D\Phi] \quad ,$$

(4.15) $$\frac{\delta F}{\delta \Phi} = \frac{\delta M}{\delta \Phi}\{\mathrm{Vec}[KGC'A(\Sigma^{-1} - W)ACGKM(B\Phi B' + \Theta)$$

$$+ (B\Phi B' + \Theta)MKGC'A(\Sigma^{-1} - W)ACGK]\}$$

$$+ L\ \mathrm{Vec}B'MKGCA(\Sigma^{-1} - W)ACKGM \quad ,$$

where

(4.16) $$\frac{\delta M}{\delta \theta} = L_{r^2}[\Phi^{-1} \otimes \Phi^{-1} \quad 0 \quad 0 \quad 0] \quad ,$$

(4.17) $$\frac{\delta F}{\delta \theta} = \Lambda_2'(\Sigma^{21} - W_{21})\Delta_p\Lambda_2(D\Phi D' + \theta)\Delta_r D + \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2(D\Phi D' + \theta)$$

$$+ D'\Delta_r(D\Phi D' + \theta)\Lambda_2'\Delta_p(\Sigma^{12} - W_{12})\Lambda_2$$

$$+ (D\Phi D' + \theta)\Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 + \Lambda_2'(\Sigma^{22} - W_{22})\Lambda_2 \quad ,$$

$$(4.18) \qquad \frac{\delta F}{\delta \psi_{11\ \mathrm{diag}}} = (\Sigma^{11} - W_{11})_{\mathrm{diag}} \quad ,$$

$$(4.19) \qquad \frac{\delta F}{\delta \psi_{12\ \mathrm{diag}}} = 2(\Sigma^{12} - W_{12})_{\mathrm{diag}} \quad ,$$

and

$$(4.20) \qquad \frac{\delta F}{\delta \psi_{22\ \mathrm{diag}}} = (\Sigma^{22} - W_{22})_{\mathrm{diag}} \quad .$$

In order to test the hypothesis

$$(4.21) \qquad H_o^{(4)} : \Lambda_1 = \Lambda_2(D\Phi D' + \theta)\Phi^{-1} \quad ,$$

against

$$(4.22) \qquad H_A^{(4)} : \Lambda_1 \neq \Lambda_2(D\Phi D' + \theta)\Phi^{-1} \quad ,$$

the model given by Equation (4.10) should be estimated with Δ_p and Δ_r constrained to be identity matrices. Denote the value of Equation (3.3), with Σ estimated as

$$\hat{\Sigma} = \hat{C}\hat{G}\hat{M}(\hat{B}\hat{\Phi}\hat{B}' + \hat{\Theta})\hat{M}\hat{G}\hat{C} + S_1 * \hat{\Psi} \quad ,$$

by $F_{min}(\Lambda_c)$. Then the likelihood ratio criterion for testing the hypothesis given by Equation (4.21) is

$$-2\log\lambda = n[F_{min}(\Lambda_c) - F_{min}] \quad ,$$

and is distributed, for large n, as χ^2 with $pr - 2r^2$ degrees of freedom.

In order to test the hypothesis

$$(4.23) \qquad H_o^{(5)} : \Lambda_1 = \Delta_p \Lambda_2 (D\Phi D' + \theta) \Delta_r \Phi^{-1} \quad ,$$

against the alternative

$$(4.24) \qquad H_A^{(5)} : \Lambda_1 \neq \Delta_p \Lambda_2 (D\Phi D' + \theta) \Delta_r \Phi^{-1} \quad ,$$

the model given by Equation (4.10) should be estimated with the diagonal elements of the matrices Δ_p and Δ_r free to vary. Denote the value of (3.3) with $\hat{\Sigma}$ estimated in this fashion by $F_{min}(\Lambda_r)$. Then the likelihood ratio criterion for testing the hypothesis given by Equation (4.23) is

$$-2\log\lambda = n[F_{min}(\Lambda_r) - F_{min}] \quad ,$$

and is distributed as χ^2 with $pr - 2r^2 - r - p$ degrees of freedom.

The next hypothesis of interest is the hypothesis of invariant factor scores. Swaminathan (1972) has pointed out that it is not possible to test this hypothesis alone. Rather, the hypothesis that the factor scores are equal or differ by a non-zero constant must be tested. The latter hypothesis implies $D = I$ and $\theta = 0$, and so we test the hypothesis

$$(4.25) \qquad H_o^{(6)} : D = I, \quad \theta = 0 \quad ,$$

against the alternative

$$(4.26) \qquad H_A^{(6)} : D \neq I, \quad \theta \neq 0 \quad .$$

In order to test the hypothesis, an estimate of $\hat{\Sigma}$ is obtained after eliminating Equations (3.18b) and (3.18d) from the likelihood equations. This estimate is substituted in Equation (3.3) and if the resulting value is denoted $F_{min}(D,\theta)$ then the likelihood ratio criterion is

$$-2\log\lambda = n[F_{min}(D,\theta) - F_{min}] \quad .$$

The test statistic is distributed as χ^2 with $2p^2 - 2pr - 2p + r(r - 1)/2$ degrees of freedom.

The last two hypotheses that will be discussed concern the unique variances. The first hypothesis is that the unique variances are invariant over the two occasions. That is, the hypothesis

$$(4.27) \qquad H_o^{(7)} : \psi_{11} = \psi_{22} \quad ,$$

which is tested against the alternative

$$(4.28) \qquad H_A^{(7)} : \psi_{11} \neq \psi_{22} \quad .$$

As usual, in order to obtain the likelihood ratio criterion, Σ must be estimated under the hypothesis given by Equation (4.27). This may be done by replacing Equations (3.18e) and (3.18g) by the single equation

$$(\Sigma^{11} - W_{11})_{diag} + (\Sigma^{22} - W_{22})_{diag} = 0 \quad .$$

If the value of Equation (3.3), with $\hat{\Sigma}$ estimated under the hypothesis of invariant unique score variances, is denoted $F_{min}(\Psi)$, then

$$-2\log\lambda = n[F_{min}(\Psi) - F_{min}] \quad ,$$

is the likelihood ratio criterion and is distributed as χ^2 on p degrees of freedom.

A second hypothesis that may be of interest, is the hypothesis that the unique factor scores are invariant over occasions. This hypothesis may be expressed as

$$(4.29) \qquad H_o^{(8)} : \psi_{11} = \psi_{12} = \psi_{22} = \psi_0 \quad ,$$

and is tested against the alternative

$$(4.30) \qquad H_A^{(8)} : \psi_{11} \neq \psi_{12}, \quad \psi_{12} \neq \psi_{22} \quad .$$

In order to obtain the likelihood ratio criterion, Equations (3.18e), (3.18f) and (3.18g) should be replaced by

$$[(\Sigma^{11} - W_{11}) + (\Sigma^{12} - W_{12}) + (\Sigma^{21} - W_{21}) + (\Sigma^{22} - W_{22})]_{diag} = 0 \quad ,$$

and Σ is estimated using the resulting equations. The value of Equation (3.3) based on this estimate of Σ may be denoted $F_{min}(\psi_0)$ and the likelihood ratio criterion for testing the hypothesis is

$$-2\log\lambda = n[F_{min}(\psi_0) - F_{min}] \quad ,$$

which is distributed as χ^2 with 2p degrees of freedom.

As has been suggested by Joreskog (1971) and Swaminathan (1972, 1973), an investigator may want to test the hypotheses concerned with parametric invariance in a sequential fashion. The method of testing hypotheses sequentially is illustrated next.

We begin by testing the hypothesis of invariant factor patterns,

(4.31) $$H_0^{(9)} : \Lambda_1 = \Lambda_2 = \Lambda_0 \quad ,$$

against

(4.32) $$H_A^{(9)} : \Lambda_1 \neq \Lambda_2 \quad .$$

As was noted previously, the likelihood ratio criterion for this hypothesis is

$$-2\log\lambda = n[F_{min}(\Lambda_0) - F_{min}] \quad ,$$

which is asymptotically distributed as χ^2 with $r(p - r)$ degrees of freedom. If this hypothesis is tenable, we may proceed to test the hypothesis that the factor scores at occasion two equal the factor scores at occasion one plus a constant, i.e.,

$$(4.33) \qquad H_o^{(10)} : \Lambda_1 = \Lambda_2, \quad D = I, \quad \theta = 0 \quad ,$$

against the alternative

$$(4.34) \qquad H_A^{(10)} : \Lambda_1 = \Lambda_2 \quad .$$

The likelihood ratio statistic is given by

$$n[F_{min}(\Lambda_0,D,\theta) - F_{min}(\Lambda_0)] \quad ,$$

where $F_{min}(\Lambda_0,D,\theta)$ is the value of (3.3) when Σ is estimated with $\Lambda_1 = \Lambda_2$, $D = I$, and $\theta = 0$, and is distributed as χ^2 with $1/2(3r^2 + r)$ degrees of freedom. If the hypothesis given by (4.33) is accepted, an hypothesis that a strict form of factorial invariance exists may be tested, i.e.,

$$(4.35) \qquad H_o^{(11)} : \Lambda_1 = \Lambda_2, \quad D = I, \quad \theta = 0, \quad \psi_{11} = \psi_{22} \quad ,$$

against

$$(4.36) \qquad H_A^{(11)} : \Lambda_1 = \Lambda_2, \quad D = I, \quad \theta = 0 \quad .$$

The likelihood ratio criterion is

$$-2\log\lambda = n[F_{min}(\Lambda_0,D,\theta,\psi_0) - F_{min}(\Lambda_0,D,\theta)] \quad ,$$

where $F_{min}(\Lambda_0,D,\theta,\psi_0)$ is the minimum value of (3.3) when $\Lambda_1 = \Lambda_2$, $D = I$, $\theta = 0$, $\psi_{11} = \psi_{22}$, and is asymptotically distributed as χ^2 with p degrees of freedom.

4.3 Asymptotic Distribution Theory

Let $\underline{\Theta}$, a column vector of dimension 2pr + r + 3p at most, contain the free parameters in Λ, B, Φ, Θ and Ψ. Let $H(\underline{\Theta})$ denote the Hessian. The Information Matrix is defined as

$$\mathcal{I}(\underline{\Theta}) = E[H(\underline{\Theta})] \quad .$$

Let $\hat{\underline{\Theta}}$ be a column vector, of dimension 2pr + r + 3p at most, which has as its elements the maximum likelihood estimates of the free parameters in Λ, B, Φ, Θ and Ψ. The asymptotic variance-covariance matrix of $\hat{\underline{\Theta}}$ is

$$C(\hat{\underline{\Theta}}) = \frac{2\,\mathcal{I}^{-1}}{n} \quad ,$$

where $\mathcal{I}$ is evaluated at the maximum of the likelihood function. It is well known that asymptotically $\hat{\underline{\Theta}}$ has a multivariate normal distribution with mean $\underline{\Theta}$ and variance-covariance matrix $C(\hat{\underline{\Theta}})$, i.e.,

$$\hat{\underline{\Theta}} \sim N[\underline{\Theta}, C(\hat{\underline{\Theta}})] \quad .$$

Thus, in order to test hypotheses about a given element or group of elements in $\hat{\underline{\Theta}}$, the available statistical theory for normally distributed random variables can be invoked.

For any given parameter Θ_i with a maximum likelihood estimate $\hat{\Theta}_i$ and variance estimate $[2/n][C(\underline{\Theta})]_{ii}$, the 100(1 - α) percent confidence interval is

$$\hat{\Theta}_i - z_\alpha\{[2/n][C(\underline{\Theta})]_{ii}\} < \Theta_i < \hat{\Theta}_i + z_\alpha\{[2/n][C(\underline{\Theta})]_{ii}\} \quad . \tag{4.37}$$

Using the confidence interval it is possible to test the hypothesis that a given parameter has specified value, i.e.,

$$H_o : \Theta_i = \Theta_0 \quad ,$$

versus

$$H_A : \Theta_i \neq \Theta_0 \quad .$$

Hypotheses about a group of elements of $\underline{\Theta}$ may be tested jointly. In order to test the hypothesis

$$H_o : K\underline{\Theta} = K\underline{\Theta}_0 \quad ,$$

against

$$H_A : K\underline{\Theta} \neq K\underline{\Theta}_0 \quad ,$$

where K is a matrix of rank q, the test statistic is

$$(4.38) \qquad \frac{n}{2}\,(\hat{\underline{\Theta}} - \underline{\Theta}_0)'K'(K\;\;{}^{-1}K)^{-1}K(\hat{\underline{\Theta}} - \underline{\Theta}_0) \quad ,$$

which is distributed at χ^2 with q degrees of freedom.

In order to utilize the procedures outlined above, expressions for the expected value of the Hessian are required. These expressions are given below. In obtaining the required expressions the following results, given by Swaminathan and McDonald (1972) will be used repeatedly:

$$E(\Sigma^{-1} - W) = 0 \quad ,$$

$$E(W) = \Sigma^{-1} \quad ,$$

$$E(X \otimes\otimes A) = E(X) \otimes\otimes A \quad ,$$

and

$$E(A \otimes\otimes X) = A \otimes\otimes E(X) \quad .$$

The results

$$(4.39) \qquad E(A \circledast X) = A \circledast E(X) \quad ,$$

and

$$(4.40) \qquad E(X \otimes A) = E(X) \otimes A \quad ,$$

are also required.

By Equation (3.33)

$$\frac{\delta^2 F}{\delta\Lambda_{diag}\delta\Lambda_{diag}} = \Sigma^{-1} \circledast \Gamma\Lambda'(W - \Sigma^{-1})\Lambda\Gamma + W \circledast \Gamma\Lambda'\Sigma^{-1}\Lambda\Gamma + (\Sigma^{-1} - W)$$

$$+ E_1[\Gamma\Lambda'\Sigma^{-1} \circledast (W - \Sigma^{-1})\Lambda\Gamma + \Gamma\Lambda'\Sigma^{-1} \circledast \Sigma^{-1}\Lambda\Gamma)] \quad .$$

and by Equations (4.39) and (4.40)

$$E \{\frac{\delta^2 F}{\delta\Lambda_{diag}\delta\Lambda_{diag}}\} = \Sigma^{-1} \circledast \Gamma\Lambda'\Sigma^{-1}\Lambda\Gamma + E_1(\Gamma\Lambda'\Sigma^{-1} \circledast \Sigma^{-1}\Lambda\Gamma) \quad .$$

Using similar steps, the following results are obtained:

$$E \{\frac{\delta^2 F}{\delta D\delta\Lambda_{diag}}\} = 2P(\Lambda'\Sigma^{-1} \otimes\otimes \Phi B\Lambda'\Sigma^{-1}\Lambda\Gamma + \Phi B'\Lambda'\Sigma^{-1} \otimes\otimes \Sigma^{-1}\Lambda\Gamma)Q \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Phi\delta\Lambda_{diag}}\right\} = 2L_{r^2}(B'\Lambda'\Sigma^{-1} \otimes\!\!\otimes\ \Sigma^{-1}\Lambda\Gamma)Q \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\theta\delta\Lambda_{diag}}\right\} = 2R(\Lambda'\Sigma^{-1} \otimes\!\!\otimes\ \Lambda'\Sigma^{-1}\Lambda\Gamma)Q \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Psi\delta\Lambda_{diag}}\right\} = 2D(S)L^*(\Sigma^{-1} \otimes\!\!\otimes\ \Sigma^{-1}\Lambda\Gamma)Q \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta D\delta D}\right\} = 2P(\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\!\otimes\ \Phi B'\Lambda'\Sigma^{-1}\Lambda B\Phi$$

$$+\ \Phi B'\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\!\otimes\ \Lambda'\Sigma^{-1}\Lambda B\Phi)$$

$$E\left\{\frac{\delta^2 F}{\delta\Phi\delta D}\right\} = 2L_r(B'\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\!\otimes\ B'\Lambda'\Sigma^{-1}\Lambda B\Phi) \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\theta\delta D}\right\} = 2R(\Lambda'\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\!\otimes\ \Lambda'\Sigma^{-1}\Lambda B\Phi) \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Psi\delta D}\right\} = 2D(S_1)L^*(\Sigma^{-1}\Lambda P^{*\prime} \otimes\!\!\otimes\ \Sigma^{-1}\Lambda B\Phi) \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Phi\delta\Phi}\right\} = L_{r^2}(B'\Lambda'\Sigma^{-1}\Lambda B \otimes\!\!\otimes\ B'\Lambda'\Sigma^{-1}\Lambda B)L_{r^2} \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\theta\delta\Phi}\right\} = R(\Lambda'\Sigma^{-1}\Lambda B \otimes\!\!\otimes\ \Lambda'\Sigma^{-1}\Lambda B)L_{r^2} \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Psi\delta\Phi}\right\} = D(S_1)L^*(\Sigma^{-1}\Lambda B \otimes\!\!\otimes\ \Sigma^{-1}\Lambda B)L_{r^2} \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\theta\delta\theta}\right\} = L_{r^2}(\Lambda_2'\Sigma^{22}\Lambda_2 \otimes \Lambda_2'\Sigma^{22}\Lambda_2)L_{r^2} \quad ,$$

$$E\left\{\frac{\delta^2 F}{\delta\Psi\delta\theta}\right\} = D(S_1)L_{p^2}(\Sigma^{22}\Lambda_2 \otimes \Sigma^{22}\Lambda_2)L_{r^2} \quad ,$$

and

$$E\left\{\frac{\delta^2 F}{\delta\Psi\delta\Psi}\right\} = D(S_1)L^*(\Sigma^{-1} \otimes\otimes \Sigma^{-1})L^*D(S_1) \quad .$$

CHAPTER V

EXTENSIONS TO THE MODEL AND LIMITATIONS OF THE STUDY

5.0 Introduction

In this chapter two extensions of the model are considered. In section 5.1, the problem of investigating the similarities and differences, in the structures of covariance matrices for multivariate data, collected on two occasions, for independent groups, is discussed. The procedures for carrying out such an investigation are briefly outlined. In section 5.2, the structural equations of a longitudinal factor model for analyzing multi-response, multi-occasion data, collected on k occasions, is presented. In section 5.3, some limitations of the study are discussed. As a whole, chapter five represents a discussion of some areas for further research in longitudinal factor analysis.

5.1 Simultaneous Longitudinal Factor Analysis in Several Populations

Suppose that p tests are given to m independent groups on two occasions, and that the model given by (2.10) fits the covariance matrix for each of the groups. We are interested in studying the similarities and differences in the structures for the m populations. Joreskog (1971) has investigated the similar problem of studying the similarities and differences in the parameters of the common factor model for m populations.

For the gth group $(g=1,2,\ldots,m)$ we assume that the linear model given by Equation (2.7) holds. That is,

$$\underline{y}_g = \Lambda_g \underline{x}_g + \underline{e}_g \quad ,$$

and that the covariance matrix for $\underline{y}_g$ has the form given by (2.10), i.e.,

$$\Sigma_g = \Lambda_g(B_g \Phi_g B_g' + \theta_g)\Lambda_g' + S_1 * \Psi_g \quad .$$

We also assume that $\underline{y}_g$ follows a multinormal distribution for each g.

Let N_g be the number of individuals in the sample from the gth population, and let S_g be the within group sample variance-covariance matrix, with $n_g = N_g - 1$ degrees of freedom. The logarithm of the likelihood function for the gth group is

$$\log L_g = -\frac{n_g}{2}\,[\log|\Sigma_g| + \mathrm{Tr}(S_g\Sigma_g^{-1})] \quad .$$

Since the samples are independent

$$\log L = \sum_{g=1}^{m} \log L_g \quad ,$$

is the logarithm of the likelihood function for all samples. Maximum likelihood estimates of the free parameters in Λ_g, B_g, Φ_g, Θ_g and Ψ_g may be obtained by maximizing logL. However, Joreskog (1971) suggests it is more convenient to minimize

$$(5.1) \qquad F = \sum_{g=1}^{m} n_g[\log|\Sigma_g| + \mathrm{Tr}(S_g\Sigma_g^{-1}) - \log|S_g| - 2p] \quad .$$

In order to minimize F, the derivatives $\delta F/\delta\Lambda_{g\ \mathrm{diag}}$, $\delta F/\delta D_g$, $\delta F/\delta\Phi_g$,

$\delta F/\delta\theta_g$ and $\delta F/\delta\Psi_g$ are required. For each group, these derivatives are given by Equations (3.18). For example

$$\frac{\delta F}{\delta \Lambda_{g\ \mathrm{diag}}} = [(\Sigma_g^{-1} - W_g)\Lambda_g\Gamma_g]_{\mathrm{diag}} \quad .$$

Following the development of Joreskog (1971), we outline the tests of hypotheses concerned with equality of the parameters for the m groups.

The first hypothesis to be tested is the hypothesis of equality of the m covariance matrices, i.e.,

$$H_o^{(1)} : \Sigma_1 = \Sigma_2 = \ . \ . \ . = \Sigma_m \quad .$$

It is well known that the hypothesis may be tested using the test statistic

$$-2\log\lambda = n\log|S| - \sum_{g=1}^{m} n_g|S_g| \quad ,$$

where

$$n = \sum_{g=1}^{m} n_g \quad ,$$

and

$$S = \frac{1}{n} \sum_{g=1}^{m} n_g S_g \quad .$$

The test statistic is distributed as χ^2, in large samples, with $p(m - 1)(2p + 1)$ degrees of freedom. If the hypothesis given by (5.2) is tenable, the pooled variance-covariance matrix may be analyzed by the longitudinal factor model, and there is no need to analyze the S_g separately. If the hypothesis is not supported, the similarities and differences in the structures for Σ_g should be examined.

The first hypothesis tested is the hypothesis that the same number of factors are operating at each occasion for each group. This is the hypothesis

$$(5.3) \qquad H_o^{(2)} : r_1 = r_2 = \ldots = r_m \quad .$$

Testing the hypothesis may be accomplished by performing separate longitudinal factor analyses, using the same number of factors, for each group. The minimum of the function given by (3.3) is computed for each group. Denote these minima by $F_{g\,min}$. Each $n_g F_{g\,min}$ is distributed as χ^2, in large samples, with $2p^2 - 2pr - 2p - r$ degrees of freedom. Since the test statistics, $n_g F_{g\,min}$, are independent

$$F_{min} = \sum_{g=1}^{m} n_g F_{g\,min} \quad ,$$

is distributed as χ^2 with $m(2p^2 - 2pr - 2p - r)$ degrees of freedom. The quantity F_{min} is the minimum value of the function given by (5.1) under the hypothesis of a common number of factors, and may be used to test the hypothesis given by Equation (5.3).

If the hypothesis of a common number of factors is tenable, we may proceed to test the hypothesis that the factor patterns for the m groups are equal. That is, the hypothesis

$$(5.4) \qquad H_o^{(3)} : \Lambda_1 = \Lambda_2 = \ldots = \Lambda_m = \Lambda_0 \quad .$$

This may be accomplished as follows. Substitute the derivative

$$\frac{\delta F}{\delta \Lambda_0} = \sum_{g=1}^{m} (\Sigma_g^{-1} - \Sigma_g^{-1} S_g \Sigma_g^{-1}) \Lambda_g \quad ,$$

for $\delta F/\delta\Lambda_1$, $\delta F/\delta\Lambda_2$, ... , $\delta F/\delta\Lambda_m$. Estimate the free parameters in Λ_0; B_1, B_2, ... , B_m; Φ_1, Φ_2, ... , Φ_m; Θ_1, Θ_2, ... , Θ_m; Ψ_1, Ψ_2, ... , Ψ_m; using the resulting equations. Compute the minimum value of the function F given by Equation (5.1). Denote this value as $F_{min}(\Lambda_0)$. Then

$$-2\log\lambda = F_{min}(\Lambda_0) - F_{min} \quad ,$$

is distributed as χ^2 with $2(m - 1)pr$ degrees of freedom.

Proceeding sequentially, if the hypothesis given by (5.4) is tenable, we test the hypothesis

$$H_o^{(4)} : \Lambda_1 = \Lambda_2 = \ldots = \Lambda_m = \Lambda_0 \quad ; \quad \Psi_1 = \Psi_2 = \ldots = \Psi_m = \Psi_0 \quad .$$

In order to obtain a test of this hypothesis, the derivative

$$\frac{\delta F}{\delta \Psi_0} = D(S_1) \quad \{\sum_{g=1}^{m} [2(\Sigma_g^{-1} - W_g) - I * (\Sigma_g^{-1} - W_g)]\} \quad ,$$

is substituted for $\delta F/\delta \Psi_1$, $\delta F/\delta \Psi_2$, ... , $\delta F/\delta \Psi_m$ and the resulting equations are solved for $\hat{\Lambda}_0$; $\hat{B}_1$, $\hat{B}_2$, ... , $\hat{B}_m$; $\hat{\Phi}_1$, $\hat{\Phi}_2$, ... , $\hat{\Phi}_m$; $\hat{\Theta}_1$, $\hat{\Theta}_2$, ... , $\hat{\Theta}_m$; $\hat{\Psi}_0$. The value of the function F, given by (5.1), is then computed with

$$\hat{\Sigma}_g = \hat{\Lambda}_0(\hat{B}_g \hat{\Phi}_g \hat{B}_g' + \hat{\theta}_g)\hat{\Lambda}_0' + S_{\cdot} * \hat{\Psi}_0 \quad .$$

If we denote this value of F as $F_{min}(\Lambda_0, \Psi_0)$, the test statistic

$$-2\log\lambda = [F_{min}(\Lambda_0, \Psi_0) - F_{min}(\Lambda_0)] \quad ,$$

is distributed as χ^2 with $3(m - 1)p$ degrees of freedom.

Using similar procedures, it is possible to develop tests of the equality of the parameter matrices B_g, Φ_g, and Θ_g, for the m groups.

5.2 A k Occasion Longitudinal Factor Analysis Model

As with the two occasion model we assume that the common factor model holds for each of k occasions. That is

$$(5.6) \qquad \underline{y}_i = \Lambda_i \underline{x}_i + \underline{e}_i \quad , \quad (i=1,2,\ldots,k) \quad ,$$

where $\underline{y}_i$, $\underline{\Lambda}_i$, $\underline{x}_i$ and $\underline{e}_i$ have the same definition as in section 2.2. We also make the same assumptions concerning means and covariances as we did in section 2.2, namely

(5.7a) $$E(\underline{y}_i) = 0 \quad ,$$

(5.7b) $$E(\underline{e}_i) = 0 \quad ,$$

(5.7c) $$E(\underline{x}_i) = 0 \quad ,$$

(5.7d) $$E(\underline{x}_i\underline{e}_j') = 0 \quad , \quad (i,j=1,2,\ldots,k) \quad ,$$

and

(5.7e) $$E(\underline{e}_i\underline{e}_j') = \psi_{ij} = \psi_{ji} \quad , \quad \text{diagonal } (i,j=1,2,\ldots,k) \quad .$$

The regression model for the factor scores is

$$\underline{x}_i = D_i\underline{x}_{i-1} + \underline{d}_i \quad , \quad (i=1,2,\ldots,k) \quad .$$

The matrix D_i is an (rxr) matrix of regression weights for the prediction of the factor scores at occasion i from the factor scores at occasion i-1. The vector $\underline{d}_i$ is an (rx1) random vector of change scores. As with the two occasion model we assume that

(5.8a) $$E(\underline{d}_i) = 0 \quad ,$$

(5.8b) $$E(\underline{x}_j\underline{d}_i') = 0 \quad , \quad (j=1,2,\ldots,i-1) \quad ,$$

(5.8c) $$E(\underline{d}_i\underline{d}_j') = 0 \quad , \quad (i \neq j) \quad ,$$

and define

(5.8d) $$E(\underline{d}_i\underline{d}_i') = \theta_i \quad .$$

Assumption (5.8b) states that the vector of change scores for occasion i is orthogonal to the factor scores for all proceeding occasions. In (5.8c) we assume that the change scores for different occasions are orthogonal. If we define

(5.9) $$E(\underline{x}_1\underline{x}_1') = \Phi_{11} \quad ,$$

and utilize the previously stated assumptions we may derive general expressions for the variance of factor scores for occasion i and the covariance of factor scores for occasions i and j (j<i). The expressions, for the variance and covariance respectively, are

(5.10a) $$E(\underline{x}_i\underline{x}_i') = \Phi_{ii} \quad ,$$

$$= D_i D_{i-1} \cdots D_2 \Phi_{11} D_2' \cdots D_{i-1}' D_i' + \theta_i$$

$$+ D_i \theta_{i-1} D_i' + D_i D_{i-1} \theta_{i-2} D_{i-1}' D_i'$$

$$+ \cdots + D_i D_{i-1} \cdots D_3 \theta_2 D_3' \cdots D_{i-1}' D_i' \quad ,$$

and

(5.10b) $$E(\underline{x}_i \underline{x}_j') = \Phi_{ij} \quad ,$$

$$= D_i D_{i-1} \cdots D_2 \Phi_{11} D_2' \cdots D_j'$$

$$+ D_i D_{i-1} \cdots D_{j+1} \theta_j + D_i D_{i-1} \cdots D_j \theta_{j-1} D_j$$

$$+ \cdots + D_i D_{i-1} \cdots D_3 \theta_2 D_3 \cdots D_j \quad .$$

In particular

(5.11) $$E(\underline{x}_i \underline{x}') = D_i D_{i-1} \cdots D_2 \Phi_{11} \quad ,$$

Denoting the collection of observed scores, factor scores and unique scores by

(5.12a) $$\underline{y}' = [\underline{y}_1 \; \underline{y}_2 \cdots \underline{y}_k] \quad ,$$

(5.12b) $$\underline{x}' = [\underline{x}_1 \; \underline{x}_2 \cdots \underline{x}_k] \quad ,$$

and

(5.12c) $\underline{e}' = [\underline{e}_1 \; \underline{e}_2 \; . \; . \; . \; \underline{e}_k]$,

we may rewrite Equation (5.6) in terms of the collected vectors,

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{y}_k \end{bmatrix} = \begin{bmatrix} \Lambda_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \Lambda_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \Lambda_k \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_k \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \cdot \\ \cdot \\ \cdot \\ \underline{e}_k \end{bmatrix} .$$

Equation (5.13) may also be expressed as

(5.14) $\underline{y} = \Lambda\underline{x} + \underline{e}$.

The structural equations are derived by finding the expected values of the observed scores in terms of the expected values of the common factor scores and specific factor scores. That is

(5.15) $E(\underline{yy}') = \Sigma = \Lambda E(\underline{xx}')\Lambda' + E(\underline{ee}')$,

$$= \Lambda\Phi\Lambda' + \Psi \quad .$$

The matrices Σ, Λ, Φ and Ψ are super matrices. The matrix Σ, of inter-occasion dispersion matrices is given by

$$(5.16a) \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdot & \cdot & \cdot & \Sigma_{1k} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \Sigma_{k1} & \cdot & \cdot & \cdot & \cdot & \Sigma_{kk} \end{bmatrix},$$

Λ, the matrix of factor patterns is given by

$$(5.16b) \qquad \Lambda = \begin{bmatrix} \Lambda_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \Lambda_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \Lambda_k \end{bmatrix},$$

Φ, the matrix of inter-occasion factor score dispersion matrices, is given by

$$(5.16c) \qquad \Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdot & \cdot & \cdot & \Phi_{1k} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \Phi_{k1} & \cdot & \cdot & \cdot & \cdot & \Phi_{kk} \end{bmatrix},$$

Ψ, the matrix of inter-occasion unique score dispersion matrices, is

$$(5.16d) \qquad \Psi = \begin{bmatrix} \psi_{11} & \psi_{12} & \cdot & \cdot & \cdot & \psi_{1k} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \psi_{k1} & \cdot & \cdot & \cdot & \cdot & \psi_{kk} \end{bmatrix} .$$

The submatric element Σ_{ij} of Σ is (pxp), Λ_i of Λ is (pxr), $\Phi_{ij} = \Phi'_{ji}$ of Φ is (rxr) and $\psi_{ij} = \psi_{ji}$ is a diagonal matrix of order (pxp).

Since Φ_{ii} and Φ_{ij} have the form given by Equations (5.10a) and (5.10b), we may not simply estimate Φ. Estimation of Φ would ignore the mathematical dependence among the submatrices of Φ and would prevent us from estimating the D_i's and θ_j's of the model. In order to estimate the parameters, we express Φ as

$$(5.17) \qquad \Phi = \Delta T \Delta^{-1} \theta \Delta^{-T} T \Delta' \quad ,$$

where Δ^{-T} indicates the transpose of Δ^{-1}, and

$$(5.18a) \qquad \Delta = \begin{bmatrix} I & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & D_2 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & D_3D_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & D_kD_{k-1}\cdot\cdot D_2 \end{bmatrix} ,$$

$$(5.18b) \qquad T = \begin{bmatrix} I & 0 & . & . & . & 0 \\ I & I & 0 & . & . & 0 \\ I & I & I & 0 & . & 0 \\ . & & & & & \\ I & I & I & . & . & I \end{bmatrix} ,$$

and

$$(5.18c) \qquad \Theta = \begin{bmatrix} \Phi_{11} & 0 & . & . & . & 0 \\ 0 & \theta_2 & 0 & . & . & 0 \\ . & & & & & \\ . & & & & & \\ 0 & . & . & . & . & \theta_k \end{bmatrix} .$$

Therefore, the structural equations for the model are:

$$(5.19) \qquad \Sigma = \Lambda(\Delta T \Delta^{-1} \theta \Delta^{-T} T \Delta')\Lambda' + \Psi \quad .$$

There is a problem that may arise in estimating the parameters D_2, D_3, ... , D_k. The problem arises because products of these matrices, rather than the matrices themselves, are estimated in the model given by Equation (5.19). That is

$$\xi_2 = D_2 \quad ,$$

$$\xi_3 = D_3 D_2 \quad ,$$

$$\begin{matrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{matrix}$$

$$\xi_k = D_k D_{k-1}, \ldots , D_2 \quad ,$$

are estimated, rather than $D_2, D_3, \ldots , D_k$.

Let Ξ be a diagonal supermatrix,

$$\Xi = \begin{bmatrix} I & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & I & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \xi_2^{-1} & \cdot & \cdot & 0 \\ 0 & 0 & 0 & \xi_3^{-1} & \cdot & 0 \\ \cdot & & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \xi_k^{-1} \end{bmatrix} .$$

The problem is that the product $\hat{\Delta}\hat{\Xi}$ may not give the maximum likelihood estimates of $D_2, D_3, \ldots , D_k$. If this product does not yield the maximum likelihood estimates of $D_2, D_3, \ldots , D_k$, it will be necessary to factor Δ into

$$\Delta = \delta_k \delta_{k-1} \cdot \cdot \cdot \delta_2 \quad ,$$

where

$$\delta_k = \begin{bmatrix} I & 0 & . & . & . & 0 \\ 0 & I & . & . & . & 0 \\ . & & & & & \\ . & & & & & \\ 0 & . & . & . & . & D_k \end{bmatrix},$$

$$\delta_{k-1} = \begin{bmatrix} I & 0 & . & . & . & 0 \\ 0 & I & . & . & . & 0 \\ . & & & & & \\ . & & & & D_{k-1} & \\ 0 & . & . & . & . & D_{k-1} \end{bmatrix},$$

and

$$\delta_2 = \begin{bmatrix} I & 0 & . & . & . & 0 \\ 0 & D_2 & . & . & . & 0 \\ . & & & & & \\ . & & & & & \\ 0 & . & . & . & . & D_2 \end{bmatrix}.$$

The model then becomes

$$\Sigma = \Lambda(\delta_k\delta_{k-1}\cdots\delta_2 T\delta_2'\cdots\delta_{k-1}'\delta_k'\Theta\delta_k^{-T}\delta_{k-1}^{-T}\cdots\delta_2^{-T}T\delta_2^{-1}\cdots\delta_{k-1}^{-1}\delta_k^{-1})\Lambda' + \Psi \quad .$$

It should be noted that, using Theorem 3.1, if it can be shown that the elements of $D_2, D_3, \ldots, D_k$, are one to one transformations of the elements of $\xi_2, \xi_3, \ldots, \xi_k$, then the product $\hat{\Delta}\hat{\Xi}$ does give the maximum likelihood estimates of $D_2, D_3, \ldots, D_k$.

5.3 Limitations of the Dissertation

In this dissertation, a longitudinal factor model that resolves many of the problems associated with previous longitudinal factor models was presented. One of the major problems still remaining is to complete develcpment of k occasion longitudinal factor model. Some progress was made on this problem in section 5.2, where the structural equations for a k occasion model was presented.

A second problem that needs to be more adequately investigated is the problem of existence and identification of the structure. The discussion of the problem, in section 2.2, applies some fairly well known results to the problem of examining the existence and identification of the structure of the two occasion longitudinal factor model. However, the discussion concerning the comparison of the number of equations with the number of unknowns, although following traditional lines, is admittedly inadequate. The theorems concerning the identification of the structure provide a better basis for investigating the identification problem, but are not without difficulties. First, the theorems only provide sufficient conditions for identification of the structure. Secondly, the results of the theorems are conditional upon the identification of ψ_{11} and ψ_{22}. No progress was made on the difficult problem of determining the conditions for the identification of these matrices.

The results given in chapters three and four provide the potential for estimating the model and testing hypotheses about the parameters. In order to realize this potential, a computer program is required. The failure to develop a program, that can carry out parameter estimation and the hypothesis testing is perhaps the most important limitation of the study. However, it should be noted that the ACOVS program developed by Joreskog, Gruvaeus, and van Thillo (1970) conceivably can be used to estimate the model. Another program that might be used for this purpose was developed by Gruvaeus and Joreskog (1970). At any rate, the problem of developing a program would seem to be a problem of adapting existing computer routines for the minimization of functions of several variables.

Associated with the failure to provide a computer program, is the lack of worked out examples of the procedures developed in the dissertation. In particular, an example would have helped clarify the hypotheses that were discussed in sections 4.1, 4.2 and 4.3.

The limitations of the dissertation discussed above and the problem of simultaneous longitudinal factor analysis in several populations represent potential areas for further research in longitudinal factor analysis. In addition, the problems of estimating factor scores, of estimating factor change scores, and of estimating the correlation of a variable with the factor change scores, as well as a number of other problems that may be extrapolated from the areas of factor analysis and the measurement of change, remain to be solved. Thus, we must conclude that longitudinal factor analysis remains a fertile area for methodological research.

References

Algina, J., and Swaminathan, H. Statistical procedures for the analysis of quasi-experimental time series designs. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975.

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Anderson, T. W., and Rubin, H. Statistical inference in factor analysis. Proceedings of the third Berkeley symposium on mathematical statistics and probability. Berkeley: University of California Press, 1956.

Bock, R. D. Multivariate analysis of variance of repeated measures. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Skokie, Illinois: Rand McNally, 1966.

Corballis, M. C. A factor model for analyzing change. British Journal of Mathematical and Statistical Psychology, 1973, 26, 90-97.

Corballis, M. C., and Traub, R. E. Longitudinal factor analysis. Psychometrika, 1970, 35, 79-98.

Fisher, F. M. The identification problem in econometrics. New York: McGraw Hill, 1966.

Glass, G. V., Willson, V. L. and Gottman, J. H. Design and analysis of time-series experiments. Boulder, Colorado: Laboratory of Educational Research, 1972.

Grizzle, J., and Allen, D. M. Analysis of growth and dose response curves. Biometrics, 1969, 25, 357-381.

Gruvaeus, G. T., and Joreskog, K. G. A computer program for minimizing a function of several variables. Research Bulletin 70-14. Princeton, New Jersey: Educational Testing Service, 1970.

Harris, C. W. Canonical factor models for the description of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Joreskog, K. G. Factoring the multitest-multioccasion correlation matrix. Research Bulletin 69-42. Princeton, New Jersey: Educational Testing Service, 1969.

Joreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-252.

Joreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 1970, 35, 111-128.

Joreskog, K. G., Gruvaeus, G. T., and can Thillo, M. A general computer program for analysis of covariance structures. Research Bulletin 70-15. Princeton, New Jersey: Educational Testing Service, 1970.

Koopmans, T. C., and Riersol, O. The identification of structural characteristics. Annals of Mathematical Statistics, 1950, 21, 165-181.

Lawley, D. N., and Maxwell, A. E. Factor analysis as a statistical method. (2nd Ed.) New York: American Elsevier Publishing Company, 1971.

Lord, F. M. Further problems in the measurement of growth. Educational and Psychological Measurement, 1958, 49, 437-451.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison: The University of Wisconsin Press, 1963.

Lord, F. M. A paradox in the interpretation of group comparisons. Psychological Bulletin, 1967, 68, 304-305.

Lord, F. M. Statistical adjustments when comparing pre-existing groups. Psychological Bulletin, 1969, 72, 336-337.

McDonald, R. P., and Swaminathan, H. Structural analysis of dispersion matrices based on a very general model with a rapidly convergent procedure for the estimation of parameters. Publication of the Department of Measurement and Evaluation: The Ontario Institute for Studies in Education, 1972.

McDonald, R. P., and Swaminathan, H. A simple matrix calculus with applications to multivariate analysis. General Systems, 1973, 18, 37-54.

Nesselroade, J. R. Note on the "Longitudinal Factor Analysis" model. Psychometrika, 1972, 37, 189-201.

O'Conner, E. F. Extending classical test theory to the measurement of change. Review of Educational Research, 1972, 42, 73-97.

Pothoff, R. F., and Roy, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 1964, 51, 313-326.

Rao, C. R. Some problems involving linear hypotheses in multivariate analysis. Biometrika, 1959, 46, 49-58.

Rao, C. R. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, 1965, 52, 447-458.

Rao, C. R. Covariance adjustment and related problems in multivariate analysis. In P. R. Krishniah (Ed.), Multivariate analysis II. New York: Academic Press, 1966.

Rao, C. R. Least squares theory using an estimated dispersion matrix and its application to the measurement of signals. Proceedings of the fifth Berkeley symposium on mathematical statistics. Berkeley: University of California Press, 1967.

Rao, C. R. Linear statistical inference and its applications. (2nd Ed.) New York: Wiley, 1975.

Riersol, O. On the identifiability of parameters in Thurstone's multiple factor analysis. Psychometrika, 1950, 15, 121-149.

Swaminathan, H. Estimation and hypothesis testing in longitudinal factor analysis. Paper presented at the Psychometric Society Meeting, Spring, 1972.

Swaminathan, H. Recent developments on the problem of factor analyzing longitudinal data. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1973.

Timm, N. H. Multivariate profile analysis of split-split plot designs and growth curve analysis of multivariate repeated measures designs. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1974.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Werts, C. E., and Linn, R. L. A general linear model for studying growth. Psychological Bulletin, 1970, 73, 17-22.

Werts, C. E., Joreskog, K. G., and Linn, R. L. A multitrait-multimethod model for studying growth. Educational and Psychological Measurement, 1972, 32, 655-678.

A P P E N D I X

A. SELECTED RESULTS IN DIFFERENTIAL CALCULUS OF MATRIX FUNCTIONS

A.0 Introduction

For the reader's convenience, we present some results in differential calculus of matrix functions. All of these results are due to McDonald and Swaminathan (1973), but are given here to clarify the process of obtaining the partial derivatives of matrix functions given in chapter three. In section A.1, we introduce the definition of a matrix derivative to be adopted, and list for later reference the commonly occurring matrix derivatives that are irreducible in the sense that appeal to scalar calculus is necessary to obtain them. In section A.2, the treatment of partitioned matrices is outlined.

In the sequel, we use the following notation:

I_n, the identity matrix of order (nxn);

X', the transpose of X;

X^{-1}, the inverse of nonsingular X;

X^{-T}, the inverse transpose of nonsingular X:

$A \otimes B$, the (right) Kronecker (or direct) product of matrices A and B, that is, $A \quad B = [a_{jk}B]$;

$A * B$, the Hadamard (elementwise) product of matrices A and B of the same order, that is, $A * B = [a_{jk}b_{jk}]$;

X_{diag}, a (nx1) vector whose components are the diagonal elements of $x_{11}, \cdots, x_{nn}$, of (nxn) X.

A.1 The Derivative of a Matrix With Respect to a Matrix

Given a (pxq) matrix Y whose elements are functions of the elements of a (mxn) matrix S, the collection of mnpq derivatives $\{\frac{\delta y_{\alpha\beta}}{\delta x_{\gamma\partial}}\}$ may be arranged as a (mnxpq) matrix to be denoted by $\delta Y/\delta X$. If $\underline{x}_s$ is the sth row of X, and $\underline{y}_r$ is the rth row of Y, we write

$$\frac{\delta \underline{y}_r}{\delta \underline{x}_s'} = \begin{bmatrix} \frac{\delta y_{r1}}{\delta x_{s1}} & \frac{\delta y_{r2}}{\delta x_{s1}} & \cdot\cdot & \frac{\delta y_{rq}}{\delta x_{s1}} \\ \frac{\delta y_{r1}}{\delta x_{s2}} & \cdot\cdot & \cdot\cdot & \frac{\delta y_{rq}}{\delta x_{s2}} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \frac{\delta y_{r1}}{\delta x_{sn}} & \cdot\cdot & \cdot\cdot & \frac{\delta y_{rq}}{\delta x_{sn}} \end{bmatrix},$$

and we define $\delta Y/\delta X$ by

$$\frac{\delta Y}{\delta X} = \begin{bmatrix} \frac{\delta \underline{y}_1}{\delta \underline{x}_1'} & \cdot & \frac{\delta \underline{y}_r}{\delta \underline{x}_1'} & \cdot & \frac{\delta \underline{y}_p}{\delta \underline{x}_1'} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\delta \underline{y}_1}{\delta \underline{x}_s'} & \cdot & \frac{\delta \underline{y}_r}{\delta \underline{x}_s'} & \cdot & \frac{\delta \underline{y}_p}{\delta \underline{x}_s'} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\delta \underline{y}_1}{\underline{x}_m'} & \cdot & \cdot & \cdot & \frac{\delta \underline{y}_p}{\delta \underline{x}_m'} \end{bmatrix}.$$

That is, the columns of $\delta Y/\delta X$ are ordered by the elements of Y written row by row to form a (1xpq) row vector, and the rows are ordered by the elements of X written row by row then transposed to form a (mnx1) column vector.

We will call a matrix derivative irreducible if it can only be written down as the collection of mnpq scalar derivatives, and reducible if it can be written as a function of matrices and one or more distinct irreducible matrix derivatives. The theorems given below enable the expression of a reducible derivative as a function of matrices and irreducible derivatives. First we give expressions for the irreducible derivatives commonly encountered in applications.

<u>Definition</u>: The matrix X is mathematically independent and variable (m.i.v.) if no element of X is constant, and no two or more elements of X are functionally dependent.

The commonly encountered irreducible derivatives are (i) $\delta X/\delta X$, (ii) $\delta X/\delta X$, X symmetric, (iii) $\delta X'/\delta X$, (iv) $\delta \mathrm{Diag}X/\delta X$, (v) $\delta(S\ \ X)/\delta X$, S constant, (vi) $\delta X/\ \mathrm{Diag}X$, (vii) $\delta X/\delta X_{diag}$.

Case (i) If X is m.i.v., of order (mxn), then

(A.1) $$\frac{\delta X}{\delta X} = I_{mn} = I_m \otimes I_n \quad .$$

Case (ii) If X is (nxn) symmetric, then

(A.2) $$\frac{\delta X}{\delta X} = L_{n^2} \quad ,$$

say, when L_{n^2} of order $(n^2 x n^2)$ is a matrix of unities and zeros with some rows and columns forming identical pairs. Let x_{jk} be the general element of the function matrix, and $x_{j'k'}$ be the general element of the argument matrix. Then the general element of L_{n^2}, ℓ_{gh}, is equal to unity if either $j = j'$ and $k = k'$, or $j = k'$ and $k = j'$, with

$$\text{(A.3)} \qquad \begin{aligned} g &= n(j - 1) + k \quad , \\ h &= n(j' - 1) + k' \quad , \quad 0 < j, j', k, k' \le n \quad , \end{aligned}$$

and is zero otherwise.

Case (iii) If X is (mxn), then

$$\text{(A.4)} \qquad \frac{\delta X'}{\delta X} = E_{mn} \quad ,$$

say, where E_{mn} is a permutation matrix of I_{mn}. With the notation indicated for Case (ii), the general element of E_{mn}, e_{gh}, is equal to unity if $j = k'$ and $k = j'$, with

$$\text{(A.5)} \qquad \begin{aligned} g &= n(j - 1) + k \quad , \quad 0 < j \le m \quad , \quad 0 < k \le n \quad , \\ h &= m(j' - 1) + k' \quad , \quad 0 < j' \le n \quad , \quad 0 < k' \le m \quad , \end{aligned}$$

and is zero otherwise.

Case (iv) If X is (nxn), then

(A.6) $$\frac{\delta \text{Diag}X}{\delta X} = J_{n^2} \quad ,$$

say, of order $(n^2 x n^2)$, where, with the notation of Case (ii), the (g,h)th element of J_{n^2} is unity if $g = h = r + (r - 1)n$, $r = 1, 2, \ldots, n$, and is zero otherwise.

Case (v) If X is (mxn), m.i.v. and S is a matrix of constants, also (mxn), then

(A.7) $$\frac{\delta (S * X)}{\delta X} = D(S) \quad ,$$

say, where D(S) is a (mnxmn) diagonal matrix whose diagonal elements d_{gg} are given by

(A.8) $$d_{gg} = s_{jk} \quad ,$$

where

(A.9) $$g = n(j - 1) + k \quad , \quad j = 1, \ldots, m; \quad k = 1, \ldots, n \quad .$$

Case (vi) If X is (nxn), we write

(A.10) $$\frac{\delta X}{\delta \text{Diag}X} = J_{n^2} \quad ,$$

of order $(n^2 x n^2)$, where $\ddot{J}_{n^2}$ is defined as for Case (iv).

Case (vii) If X is (nxn), we write

$$\text{(A.11)} \qquad \frac{\delta X}{\delta X_{diag}} = K \quad ,$$

of order $(n x n^2)$, where the (g,h)th element of K is unity if $h = n(g - 1) + g$, with $0 < g \leq n^2$, and zero otherwise.

A.2 The Treatment of Partitioned Matrices

Let $Y = [Y_{ij}]$ be a (pxq) matrix, partitioned into $(k_1 x k_2)$ submatrices, Y_{ij}, of order $(p_i x q_j)$, $i = 1, \ldots, k_1$: $j = 1, \ldots, k_2$, so that

$$\sum_{i=1}^{k_1} p_i = p \quad \text{and} \quad \sum_{j=1}^{k_2} q_j = q \quad .$$

Similarly, let $X = [X_{rs}]$ be a (mxn) matrix, partitioned into $(\ell_1 x \ell_2)$ submatrices X_{rs}, of order $(m_r x n_s)$ $r = 1, \ldots, \ell_1$; $s = 1, \ldots, \ell_2$, so that

$$\sum_{r=1}^{\ell_1} m_r = m \quad \text{and} \quad \sum_{s=1}^{\ell_2} n_s = n \quad .$$

If we write each submatrix derivative $\delta Y_{ij}/\delta X_{rs}$, of order $(m_r n_s x p_i q_j)$, $i = 1, \ldots, k_1$; $j = 1, \ldots, k_2$; $r = 1, \ldots, \ell_1$; $s = 1, \ldots, \ell_2$, according to the definition of a matrix derivative given in section A.1, we may define the partitioned matrix derivative $\delta Y/\delta X$ by

$$\text{(A.12)} \qquad \frac{\delta Y}{\delta X} = \begin{bmatrix} \frac{\delta Y_{11}}{\delta X_{11}} & \frac{\delta Y_{12}}{\delta X_{11}} & \cdot\cdot & \frac{\delta Y_{k_1k_2}}{\delta X_{11}} \\ \frac{\delta Y_{11}}{\delta X_{12}} & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \frac{\delta Y_{11}}{\delta X_{\ell_1\ell_2}} & \cdot\cdot & \cdot\cdot & \frac{\delta Y_{k_1k_2}}{X_{\ell_1\ell_2}} \end{bmatrix} .$$

It is readily verified that, in general, the partitioned derivative (A.12) is distinct from its non-partitioned counterpart obtained simply by forming the derivative of the entire matrix Y with respect to the entire matrix X, in the ordering of the scalar derivatives in the two-dimensional array. Clearly, also, the ordering of the derivatives would vary under arbitrary re-partitionings of the matrices. In some applications we have a fixed, prescribed partitioning of the matrices Y and X, as when previously defined matrices are adjoined to form a supermatrix. For such cases it may be convenient to obtain the derivatives in the order given by (A.12). The results of the previous sections can be modified to yield this order.

It is unnecessary to treat these modifications exhaustively here, as they are easily obtained when needed. It is, indeed, inconvenient to provide an exhaustive treatment, primarily because there is a much greater variety of commonly desired irreducible forms. These can always be obtained as needed by direct application of (A.12). [For example, if X is symmetric, and regularly partitioned into (2x2) submatrices, we find on application of (A.12) that

$$\text{(A.13)} \qquad \frac{\delta X}{\delta X} = \begin{bmatrix} L_{n^2} & 0 & 0 & 0 \\ 0 & I_{n^2} & E_{n^2} & 0 \\ 0 & E'_{n^2} & I_{n^2} & 0 \\ 0 & 0 & 0 & L_{n^2} \end{bmatrix} ,$$

where $L_{n^2} = \dfrac{\delta X_{ii}}{\delta X_{ii}}$ as by (A.2), and $E_{n^2} = \dfrac{\delta X_{21}}{\delta X_{12}} = \dfrac{\delta X'_{12}}{\delta X_{12}}$ as by (A.4)].

It should, therefore, suffice to develop only the basic principles and major theorems for partitioned matrices.

Definition: If A is partitioned into submatrices A_{ij}, then

$$\text{(A.14)} \qquad \mathrm{Vec}A = \begin{bmatrix} \mathrm{Vec}(A_{11}) \\ \mathrm{Vec}(A_{12}) \\ \cdot \\ \cdot \\ \cdot \\ \mathrm{Vec}(A_{k_1 k_2}) \end{bmatrix} ,$$

where $Vec(A_{ij})$, $i = 1, \ldots, k_1$; $j = 1, \ldots, k_2$, is defined as

$$Vec(A_{ij}) = \begin{bmatrix} \underline{a}'_{1ij} \\ \cdot \\ \cdot \\ \cdot \\ \underline{a}'_{mij} \end{bmatrix}$$

where $\underline{a}_{1ij}$ is the first row of A_{ij} and $\underline{a}_{mij}$ is the last row of A_{ij}.

<u>Definition</u>: If A is partitioned into (kxk) submatrices, then

$$\text{(A.15)} \qquad VecA_{diag} = \begin{bmatrix} Vec(A_{11}) \\ Vec(A_{22}) \\ \cdot \\ \cdot \\ \cdot \\ Vec(A_{kk}) \end{bmatrix} .$$

<u>Definition</u>: Let $A = [A_{ij}]$ be a (pxq) matrix, partitioned into $(k_1 x k_2)$ submatrices A_{ij}, of order $(p_i x q_j)$ $i = 1, \ldots, k_1$: $j = 1, \ldots, k_2$, and let $B = [B_{rs}]$ be a (mxn) matrix, partitioned into $(\ell_1 x \ell_2)$ submatrices B_{rs}, of order $(m_r x n_s)$, $r = 1, \ldots, \ell_1$; $s = 1, \ldots, \ell_2$. The (right) Double Kronecker Product C, of A and B, of order (mpxnq), is denoted by $A \otimes\otimes B$, and defined by

$$\text{(A.16)} \qquad C = A \otimes\otimes B = [A_{ij} \otimes B_{rs}] ,$$

where the rule for ordering the submatrices $A_{ij} \otimes B_{rs}$, of order $(p_i m_r x q_j n_s)$, to form the entire matrix C, is that for ordering the scalar elements in the (right) ordinary Kronecker product $A \otimes B = [a_{ij}B]$, where $B = [b_{rs}]$, with submatrices A_{ij}, B_{rs} replacing scalar elements a_{ij}, b_{rs}, and the Kronecker product replacing the scalar product. Note that the Double Kronecker Product consistently reduces to the ordinary product in case either every submatrix of A and B is (1x1), or there is just one submatrix which is the entire matrix A, and one which is the entire matrix B. (This is in contrast to the Hadamard-Kronecker Product defined next).

It is readily verified that the Double Kronecker Product enjoys the following properties:

(A.17) $$(A \otimes\otimes B) \otimes\otimes C = A \otimes\otimes (B \otimes\otimes C) \quad ,$$

(A.18) $$A \otimes\otimes (B + C) = A \otimes\otimes B + A \otimes\otimes C \quad ,$$

(A.19) $$(A \otimes\otimes B)' = A' \otimes\otimes B' \quad ,$$

(A.20) $$(A \otimes\otimes B)(C \otimes\otimes D) = (AC \otimes\otimes BD) \quad ,$$

and

(A.21) $$(A \otimes\otimes B)^{-1} = (A^{-1} \otimes\otimes B^{-1}) \quad .$$

<u>Definition</u>: Let A be a (pxq) matrix partitioned into $(k \times \ell)$ submatrices A_{ij} of order $(p_i x q_j)$, and B be a (mxn) matrix also partitioned into $(k \times \ell)$ submatrices B_{ij}, of order $(m_i x n_j)$, so that

$$\sum_{i=1}^{k} p_i = p, \quad \sum_{j=1}^{\ell} q_j = q, \quad \sum_{i=1}^{k} m_i = m, \quad \sum_{j=1}^{\ell} n_j = n \quad .$$

Then the Hadamard-Kronecker Product, C, of A and B, of order (gxh), where

$$g = \sum_{i=1}^{k} p_i m_i ; \quad h = \sum_{j=1}^{\ell} q_j n_j \quad ,$$

is denoted by $A \circledast B$ and defined by

$$\text{(A.22)} \qquad C = A \circledast B = [A_{ij} \otimes B_{ij}] \quad ,$$

that is, the (i,j)th submatrix, of order $(p_i m_i x q_j n_j)$ of C, is the Kronecker product $A_{ij} \otimes B_{ij}$. Note that, in general, $A \circledast B \neq B \circledast A$. Note also that in the case where we say that every submatrix of A and B is (1x1), the Hadamard-Kronecker product reduces to the Hadamard product, whereas in the case where we say that there is just one submatrix A and just one submatrix B, it reduces to the ordinary Kronecker product.

<u>Theorem A.1</u>: If the elements of a (rxs) matrix Z are functions of elements of a (pxq) matrix Y which in turn are functions of elements of a

(mxn) matrix X, then

$$\text{(A.23)} \qquad \frac{\delta Z}{\delta X} = \frac{\delta Y}{\delta X}\ \frac{\delta Z}{\delta Y} \quad ,$$

for an appropriate prescribed partitioning of the matrices.

Theorem A.2: If A, Y, B, are matrices partitioned conformably for multiplication to yield the product Z = A Y B, then

$$\text{(A.24)} \qquad \frac{\delta Z}{\delta X} = \frac{\delta Y}{\delta X} (A' \otimes\otimes B) \quad ,$$

where X is also, in general, a partitioned matrix.

Theorem A.3: If the elements of the (pxr) matrix U and of the (rxq) matrix V are functions of the elements of the (mxn) partitioned matrix X, and U, V, are partitioned conformably for multiplication,

$$\text{(A.25)} \qquad \frac{\delta}{\delta X} U\ V = \frac{\delta U}{\delta X} (I^{(u)} \otimes\otimes V) + \frac{\delta V}{\delta X} (U' \otimes\otimes I^{(v)}) \quad ,$$

where the identity matrices $I^{(u)}$ and $I^{(v)}$ are regularly partitioned, and, respectively, of orders equal to the row-order of U and the column-order of V, and partitioned conformably for premultiplication of U and postmultiplication of V.

Theorem A.4: If

$$\frac{\delta Y}{\delta X} = F \otimes\!\otimes G \quad , \tag{A.26}$$

where F, G, are both partitioned into (kxℓ) submatrices, then

$$\frac{\delta Y_{diag}}{\delta X_{diag}} = F \circledast G \quad . \tag{A.27}$$

Theorem A.5: If A is (nxm), X is (mxn), and A is partitioned identically as X', then

$$\frac{\delta X'}{\delta X} = \text{Vec}A = \text{Vec}A' \quad . \tag{A.28}$$

Theorem A.6: If X, (nxn), is symmetric and partitioned such that $X_{ii} = X'_{ii}$, and $X_{ij} = X'_{ij}$, and if A is partitioned identically as X, then

$$\frac{\delta X}{\delta X} \text{Vec}A = \text{Vec}(A + A' - I * A) \quad . \tag{A.29}$$

Theorem A.7:

$$(A \otimes\!\otimes B)\text{Vec}X = \text{Vec}(A\ X\ B') \quad . \tag{A.30}$$